

Molekulare Ähnlichkeiten und deren biologische Bedeutung

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Theoretische Biologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
Humboldt-Universität zu Berlin

von
Dipl.-Biochem. Stephan Lorenzen
geboren am 23.03.1976 in Flensburg

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jürgen Mlynek

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Thomas Buckhout, Ph. D.

Gutachter:

1. Prof. Dr. Cornelius Frömmel
2. Prof. Dr. Andreas Herrmann
3. Prof. Dr. Dietmar Schomburg

Tag der mündlichen Prüfung: 6. Januar 2006

Wodurch die Menschen in die größten
Ungewissheiten und Zweifel kommen,
das sind die Ähnlichkeiten der Dinge.

Lü Bu We
chinesischer Kaufmann,
Politiker und Philosoph
ca. 300 v.Chr. – ca. 235 v.Chr.

Abstract

This work is dealing with the biological impact of similarities between chemical structures, protein sequence motifs and local sequence surrounding as well as global sequence similarity. All four aspects are analyzed by computational methods.

The first part is dealing with chemical similarities. Based on a recently published set of prion protein misfolding inhibitors, a data base of approved drugs has been screened for compounds with chemical and structural similarities to these substances. 16 drugs are proposed as new potential inhibitors of prion protein aggregation.

The next part addresses similarities of sequence motifs which mediate the interaction with the peroxisomal membrane protein import receptor Pex19. In cooperation with an experimental group, the binding site could be characterized, and amino acid preferences of the different positions of the motif have been determined. The binding motif is a probably helical region of target proteins bearing branched aliphatic and basic residues. A position specific scoring matrix for the prediction of Pex19 binding sites could be generated and validated.

The relation between local sequence similarity and prolyl bond conformation is examined in the third part. Amino acid preferences of neighboring residues differ between cis and trans prolyl residues, and both species show different amino acid exchange patterns upon mutation. In contrast to local sequence similarity, overall sequence similarity between proteins as low as 20% is a much better indicator for the occurrence of cis prolyl bonds.

The last part focuses on inverse sequence similarity between proteins which occurs far more often than expected by chance. Proteins from a nonredundant data set have been aligned in parallel and antiparallel, and structural similarities between the detected protein pairs have been examined. It could be shown that, with the exception of short secondary structural elements, inverse sequence similarity does not imply structural similarity.

Keywords:

sequence analysis, protein folding, cis proline, binding site

Zusammenfassung

Die vorliegende Arbeit untersucht mit bioinformatischen Methoden die biologische Bedeutung von Ähnlichkeiten in Kleinstrukturen und peptidischen Sequenzmotiven sowie lokaler und globaler Sequenzähnlichkeit.

Der erste Teil der Arbeit behandelt chemische Ähnlichkeiten. Ausgehend von bekannten Inhibitoren der Fehlfaltung des Prionproteins wurde eine Datenbank pharmakologischer Wirkstoffe nach chemisch und strukturell ähnlichen Substanzen durchsucht und 16 Substanzen als neue potentielle Inhibitoren der Fehlfaltung vorgeschlagen.

Der nächste Teil untersucht Ähnlichkeiten in Sequenzmotiven, die eine Interaktion mit Pex19, dem Importrezeptor für peroxisomale Membranproteine, vermitteln. In Zusammenarbeit mit einer experimentellen Arbeitsgruppe konnte die Bindestelle charakterisiert und Präferenzen für bestimmte Aminosäuren herausgearbeitet werden. Das Bindemotiv ist eine vermutlich helikale Region mit verzweigtkettigen aliphatischen und basischen Aminosäuren. Aus experimentellen Daten konnte eine positionsabhängige Vorhersagematrix erstellt und validiert werden.

Die Beziehung zwischen lokalen Sequenzähnlichkeiten und der Konformation von Prolylbindungen in Proteinen ist Thema des dritten Teils. Die Aminosäurepräferenzen in der Nachbarschaft von cis- und trans-Prolylresten unterscheiden sich, und beide zeigen unterschiedliche Austauschpräferenzen bei Mutationen. Im Gegensatz zu lokaler Sequenzähnlichkeit ist eine globale Sequenzähnlichkeit von nur 20% ein wesentlich besserer Indikator für das Auftreten von cis-Prolylbindungen.

Der letzte Teil befaßt sich mit inverser Sequenzähnlichkeit zwischen Proteinen, die wesentlich öfter auftritt als erwartet. Proteine aus einem nicht-redundanten Datensatz wurden gleich- und gegenläufig aligniert und strukturelle Ähnlichkeiten zwischen den aufgefundenen Proteinpaaren untersucht. Es konnte gezeigt werden, daß bis auf kurze Sekundärstruktur-Einheiten eine inverse Sequenzähnlichkeit zwischen Proteinen keine strukturelle Ähnlichkeit impliziert.

Schlagwörter:

Sequenzanalyse, Proteinfaltung, cis-Prolin, Ligandenbindung

Inhaltsverzeichnis

1	Einleitung	1
1.1	Viele Sequenzen, wenige Strukturen	1
1.2	Millionen von Kleinstrukturen	3
1.3	Aus Ähnlichkeiten lernen	4
1.4	Ähnlichkeiten zwischen Kleinstrukturen	6
1.4.1	Vorhandensein chemischer Gruppen	6
1.4.2	Dreidimensionale Struktur	7
1.5	Ähnlichkeiten zwischen Proteinen	9
1.5.1	Sequenzalignments	9
1.5.2	Sequenzmotive	12
1.5.3	Strukturvergleiche	14
1.6	Zielsetzung der Arbeit	19
2	Kleinstrukturen: Inhibition der Prionenaggregation	21
2.1	Fehlfaltung von Proteinen	21
2.2	Inhibitoren der Bildung von PrP-Sc	22
2.3	Zielsetzung	24
2.4	Ausgangspunkte für neue Leitstrukturen	24
2.5	Kleinstruktur-Datenbanken	25
2.6	2D-Struktursuche	25
2.7	3D-Struktursuche	32
2.8	16 neue potentielle Inhibitoren	33
2.9	Diskussion	33
3	Ähnlichkeiten zwischen Sequenzmotiven und ihre funktionelle Bedeutung	36
3.1	Peroxisomaler Proteinimport	36
3.1.1	Peroxisomen haben wichtige Aufgaben im Stoffwechsel	36
3.1.2	Import peroxisomaler Matrixproteine	37
3.1.3	Import peroxisomaler Membranproteine	38
3.2	Zielsetzung	38

3.3	Peptide aus Pex13 interagieren mit Pex19	40
3.4	Pattern-Suchen	44
3.5	Vorhersagematrix	45
3.6	Permutationen	52
3.7	Konservierung der Bindestelle	52
3.8	Diskussion	56
4	Lokale Sequenzähnlichkeiten	58
4.1	Vorhersagen aufgrund lokaler Sequenzumgebungen	58
4.2	Bedeutung von <i>cis</i> -Prolylbindungen	59
4.3	Zielsetzung	61
4.4	<i>cis</i> -Prolylbindungen in bekannten Proteinstrukturen	61
4.4.1	Abhängigkeit von der Auflösung	61
4.4.2	Abhängigkeit von paarweisen Sequenzidentitäten	62
4.4.3	Abhängigkeit von B-Faktoren	63
4.4.4	Datensatz	64
4.5	Vorhersage mittels Propensities	65
4.5.1	Methode	65
4.5.2	Prediktiver Wert verschiedener Positionen	67
4.6	Konservierung von Prolylresten	70
4.7	Lokale Sequenzvergleiche	72
4.8	Strukturelle Alignments	73
4.9	Sequenzalignments	74
4.10	Diskussion	77
5	Globale Sequenzähnlichkeiten	80
5.1	Sequenzähnlichkeit und Homologie	80
5.2	Hat inverse Sequenzähnlichkeit eine Bedeutung?	81
5.3	Zielsetzung	82
5.4	Die PDB als Datenquelle	83
5.5	Inverse Sequenzalignments	83
5.6	Alignments mit Zufallssequenzen	84
5.7	Generierung der Alignment-Datensätze	85
5.8	Strukturelle Überlagerungen	86
5.9	Familienzugehörigkeiten alignierter Proteine	88
5.10	Strukturelle Verwandtschaft	90
5.11	Unterschiede zwischen Alignments	92
5.11.1	Zuordnung von Sekundärstrukturen	93
5.12	Diskussion	94
6	Übergreifende zusammenfassende Betrachtung	96

Abbildungsverzeichnis

1.1	BILD-Schlagzeile	1
1.2	Funktion der Proteine im <i>E. coli</i> Genom	2
1.3	Ähnlichkeit	4
1.4	Intransitivität der Ähnlichkeit bei Strings	5
1.5	Sterol-Grundgerüst und Phenothiazine	6
1.6	Fingerprints von Kleinstrukturen	7
1.7	3D-Ähnlichkeiten zwischen Neuroleptika	9
1.8	Multiples Alignment von Prionproteinen	10
1.9	Needlamen-Wunsch-Algorithmus	11
1.10	Struktur der α -Helix	16
1.11	Struktur des β -Faltblattes	16
1.12	Torsionswinkel	17
1.13	Ramachandran-Plot	18
2.1	Struktur des Prionproteins	23
2.2	Faltungsinhibitoren des Prionproteins	26
2.3	Überlagerung von Prochlorperazin mit Opipramol	32
2.4	Potentielle Fehlfaltungsinhibitoren	34
3.1	Peptidscan von Pex13	40
3.2	Verkürzungskonstrukte aus Pex13	41
3.3	Längenanalyse des Peptids aus Pex13	42
3.4	Mutationsanalyse des Peptids aus Pex13p	43
3.5	Interaktion der Pattern-Hits mit Pex19	44
3.6	Peptidscan von Pex11	46
3.7	Mutationsanalyse der Pex19-Bindestelle in Pex11	47
3.8	Peptidscan von Pex12	50
3.9	Peptidscan von Pex25	50
3.10	Peptidscan von Pex27	51
3.11	Permutation des Pex13-Bindepeptides	53
3.12	Vorhergesagte Bindebereiche humaner Peroxine	55
3.13	Peptidscan von Pex26	56

4.1	Winkel ω in Peptidbindung	59
4.2	Anteil <i>cis</i> -Proline bei verschiedenen Auflösungen	62
4.3	Anteil <i>cis</i> -Proline bei verschiedenen Sequenzidentitäten	63
4.4	Anteil <i>cis</i> -Proline bei verschiedenen B-Faktoren	64
4.5	Vorhersagegüte unter Verwendung unterschiedlicher Sequenz- positionen	68
4.6	Vorhersagegüte mit \pm zwei Aminosäurepositionen	69
4.7	Prediktive Aussagekraft einzelner Aminosäurepositionen	69
4.8	Vorhersage mit asymmetrischen Aminosäurefenstern	70
4.9	Konservierungsgrad von Prolyl- und Nachbarresten	72
4.10	Austauschfrequenzen zu anderen Aminosäuren	73
4.11	Anteil von <i>cis</i> -Prolylresten bei verschiedenen lokalen Sequenz- identitäten	74
4.12	Konservierung von <i>cis</i> -Prolylresten in strukturellen Alignments	75
4.13	Vorhersage von <i>cis</i> -Prolinen durch Alignments	76
4.14	Wichtung der Alignments	77
5.1	Anzahl der Alignments bei verschiedenen Sequenzidentitäts- Schwellenwerten	84
5.2	Längenverteilung der Alignments	86
5.3	RMSDs der überlagerten vorwärts alignierten Proteine	87
5.4	RMSDs der überlagerten invers alignierten Proteine	87
5.5	RMSDs der überlagerten invers alignierten Proteine mit Spie- gelbildern	88
5.6	RMSDs der überlagerten alignierten Proteine aus SCOP	89
5.7	Anteil von Proteinen gleicher Familien in Alignments	90
5.8	Inverse Ähnlichkeit zwischen zwei Proteinen	91
5.9	Hydrophobe Patches in Alignments	92
5.10	Sekundärstrukturen alignierter Proteine	93

Tabellenverzeichnis

2.1	2D-Suchen in SuperDrug und MS Discovery	28
3.1	Targeting-Bereiche peroxisomaler Membranproteine	39
3.2	Pex19-Bindestellen peroxisomaler Membranproteine	39
3.3	Vorhersagematrix für Pex19-Bindung	49
3.4	Vorhergesagte Bindestellen humaner Peroxine	54
4.1	Non-Prolyl <i>cis</i> -Aminosäuren	66
4.2	Logarithmierte Propensities der einzelnen Aminosäuren	71
4.3	Superfamilien für strukturelles Alignment	75
5.1	Alignments mit Zufalls-, Vor- und Rückwärtssequenzen	85

Kapitel 1

Einleitung

1.1 Viele Sequenzen, wenige Strukturen

Die Entschlüsselung des humanen Genoms [1] ist nicht nur in der Fachpresse mit großer Spannung erwartet worden. In die Verfügbarkeit der DNA-Sequenzen wurde große Hoffnung gesetzt, was sich nicht zuletzt in der waghalsigen Behauptung der Boulevardpresse niederschlug, nach diesem Meilenstein der Forschung seien diverse Krankheiten heilbar (Abb. 1.1). Rasch machte sich allerdings Ernüchterung breit – der unmittelbare Nutzen der riesigen Flut von Daten schien eher gering.

Forscher entschlüsselten Traurigkeits-Gen

Durchbruch in der Gen-Forschung: Das Traurigkeits-Chromosom ist entschlüsselt! Deutsche und japanische Wissenschaftler hoffen, bald Depressionen, Alzheimer (Opfer: Ronald Reagan) und Parkinson (Opfer: der Papst) heilen zu können. Auch die schwere ge-	netische Krankheit „Down-Syndrom“, umgangssprachlich „Mongolismus“, und bestimmte Krebsarten werden durch dieses Chromosom verursacht. Chromosomen sind mikroskopisch kleine fadenförmige Gebilde in menschlichen Zellkernen, sie tragen unsere Erbanla-	gen, die Gene. Die Reihenfolge der Gene im Chromosom ist der Schlüssel zur Entwicklung unseres Körpers. Die Forscher fanden noch eine zweite Überraschung: Nur 40 000 Gene reichen, um einen Menschen zu bilden. Früher hatte man bis zu 140 000 Gene vermutet.
--	--	---

Abbildung 1.1: Schlagzeile der BILD-Zeitung vom 09. Mai 2000 nach der Sequenzierung des humanen Chromosoms 21. Das Blatt sieht die Heilung von Depressionen, Alzheimer, Parkinson und Down-Syndrom kurzfristig bevorstehen.

Schon vor der Sequenzierung des humanen Genoms war 1997 das Erbmaterial des Bakteriums *E. coli* [2] und im gleichen Jahr als Ergebnis der bisher größten internationalen Zusammenarbeit das erste eukaryotische Genom (die Bäckerhefe *S. cerevisiae*) publiziert worden [3]. Schon im Genom des Bakteriums ist die Funktion eines Großteils der codierten Proteine unbekannt

(Abb. 1.2), in der Hefe konnte sogar nur 43,3% der ca. 6 000 Proteine eine Funktion zugewiesen werden [4]. Selbst heute liegt der Anteil der annotierten Gene im menschlichen Genom nur bei knapp über 50% [5].

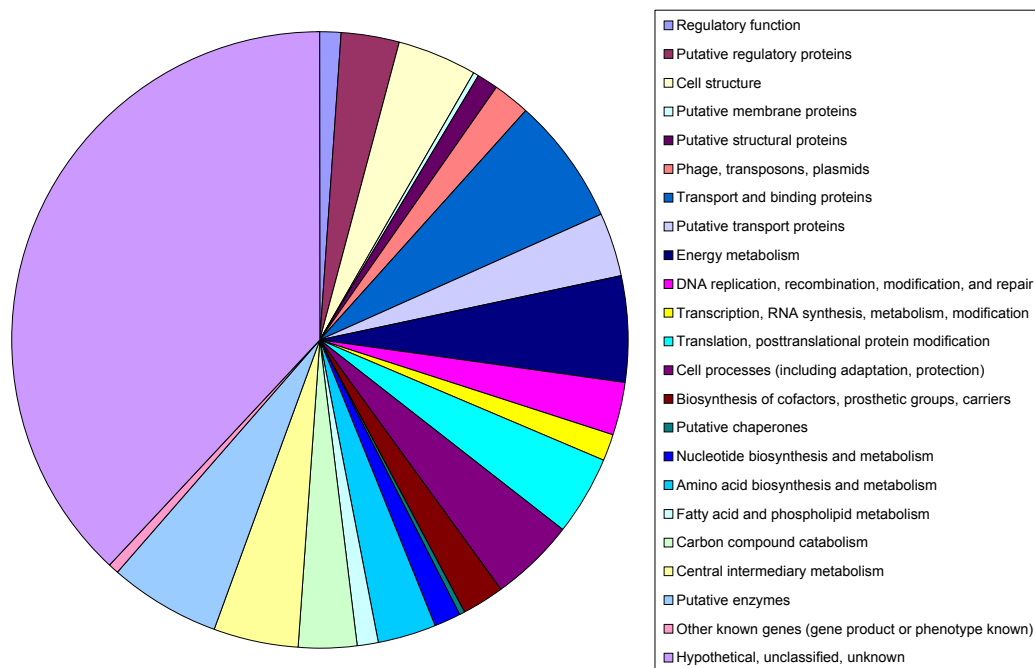


Abbildung 1.2: Funktionszuweisung der Proteine im Genom von *E. coli*. 38% der Proteine sind hypothetisch bzw. können keiner Funktion zugeordnet werden. Daten aus [2].

Eine noch wesentlich größere Lücke klappt allerdings zwischen der Zahl bekannter Sequenzen und in der *Protein Data Bank* (PDB [6]) hinterlegten Proteinstrukturen. Obwohl die Zahl bekannter Strukturen in beeindruckender Geschwindigkeit steigt und inzwischen bei ca. 30 000 liegt, ist nur ein kleiner Teil des Sequenzraumes strukturell abgedeckt – in der annotierten *SwissProt*-Datenbank [7] sind mittlerweile ca. 170 000 Sequenzen gespeichert, die nicht-redundante *RefSeq*-Datenbank des amerikanischen *National Cancer Institute* (NCI) [8] umfasst ca. 1 300 000 Sequenzen. Um diese Lücke zu schließen, wird in der *Proteinstrukturfabrik* [9], einem weltweiten *high throughput*-Projekt, systematisch versucht, eine große Zahl bislang unbekannter Proteine zu kristallisieren und ihre Struktur aufzuklären.

Die Kenntnis der dreidimensionalen Struktur eines Proteins ist eine wichtige Voraussetzung zum Verständnis seiner Wirkungsweise. So läßt sich zum Beispiel an Strukturen von Enzymen mit gebundenen Substraten oder Substratanaloga anhand der Kontakte des Substrats mit dem Protein dessen

Wirkungsmechanismus aufklären. Gut aufgelöste Proteinstrukturen erlauben auch die Suche nach spezifischen Inhibitoren, die beim *drug design* passend zur Bindestelle des Enzyms modelliert werden. Die Herangehensweise besteht dabei darin, eine dem Substrat strukturell ähnliche Struktur zu finden, die an das Enzym bindet, chemisch aber nicht umgesetzt werden kann. Vergleiche von Strukturen mit verschiedenen Inhibitoren geben Hinweise zur Verbesserung von Affinität und Spezifität, indem der Inhibitor genauer an das Enzym angepasst wird. Als Beispiel sei hier die Suche nach spezifischen Inhibitoren der HIV-Protease [10] angeführt.

Ein Verständnis des Reaktionsmechanismus eines Enzyms ermöglicht auch die gezielte Veränderung der Aktivität. Die Kristallstruktur einer reversen Transkriptase [11] ließ zum Beispiel erkennen, daß ein Phenylalanylrest des Enzyms in enger Nachbarschaft zum 2'-Kohlenstoffatom des an der Reaktion beteiligten Desoxyribonukleotids steht. Die Spezifität des Enzyms für Desoxyribonukleotide erklärt sich somit durch die sterische Behinderung bei der Bindung von Nukleotiden mit 2'-OH-Gruppe. Tatsächlich konnte die DNA-Polymeraseaktivität durch Mutation des entsprechenden Phenylalanylrestes zu Alanin in eine RNA-Polymeraseaktivität verwandelt werden [12].

Die Untersuchung der dreidimensionalen Struktur von Proteinen bietet also weit mehr Information als die bloßen Sequenzen und ist Voraussetzung für das Verständnis der Funktion eines Proteins.

1.2 Millionen von Kleinstrukturen

Ähnlich sieht die Lage bei chemischen Kleinstrukturen aus. Ständig werden von Synthesechemikern neue Substanzen hergestellt. Zur Zeit sind ca. 2 Millionen Substanzen kommerziell verfügbar [13]. Über die meisten dieser Substanzen ist allerdings wenig über ihre Wirkung bekannt. Zielsetzung der Synthese ist oft die pharmakologische Wirksamkeit gegen Krankheiten – so werden bereits bekannte Wirkstoffe in chemisch leicht veränderter Form synthetisiert und getestet.

Es ist davon auszugehen, daß sich unter den vielen Substanzen durchaus pharmakologisch interessante Wirkstoffe befinden, von deren Wirksamkeit noch nichts bekannt ist. In der Tat wurden in jüngerer Zeit viele Wirkstoffe auch gegen Krankheiten getestet, die von ihrer ursprünglichen Indikation abweichen (zur Übersicht s. [14]). Da die „Wiederentdeckung“ von etablierten Wirkstoffen auch umfangreiche klinische Tests zur Toxizität unnötig macht, ist das Auffinden von Zweitindikationen zugelassener Medikamente von großem Interesse.

Derartige Tests werden heute meist durch *high throughput screening* (HTS)

durchgeführt: Existiert ein einfacher Versuchsaufbau zum Testen der Wirksamkeit einer Substanz, werden zehntausende Substanzen getestet. Die Trefferquote ist bei derartigen Verfahren allerdings gering (in der Größenordnung von 1‰), und oft fallen hohe Kosten an. Interessant wäre demnach eine Vorauswahl geeigneter Kandidaten – zu den einzelnen Substanzen müßte eine Aussage getroffen werden, wie *wahrscheinlich* eine Aktivität ist.

1.3 Aus Ähnlichkeiten lernen

Sowohl zu Proteinen, deren Sequenz bekannt ist, als auch zu chemischen Kleinstrukturen möchte man also Aussagen treffen, ohne zeit- und kostenintensive Experimente durchführen zu müssen. Informationen über ein neues Objekt ergeben sich dabei nur durch das Aufspüren von Ähnlichkeiten zu bereits bekannten Objekten. Dabei ist der Begriff der Ähnlichkeit zwischen zwei Dingen durchaus nicht immer klar definiert.



Abbildung 1.3: Augenscheinliche Ähnlichkeit zwischen zwei Dingen ist mathematisch bisweilen schwer faßbar.

Am einfachsten sieht die Situation in der Geometrie aus: Zwei Dreiecke sind ähnlich, wenn das Verhältnis ihrer Seitenlängen zueinander und damit

auch die eingeschlossenen Winkel gleich sind. Zwei ähnliche Dreiecke lassen sich somit durch Vergrößern oder Verkleinern ineinander überführen, und es gilt das Gesetz der Transitivität: Ist A zu B ähnlich und B zu C , so ist auch A zu C ähnlich.

Im Falle der Dreiecke liegt dies schlicht daran, daß Dreiecke mit Ausnahme ihrer Seiten und Winkel keine anderen Eigenschaften besitzen. Komplizierter gestaltet sich die Situation beim Vergleich komplexerer Dinge. So ist zum Beispiel das Mädchen in Abb. 1.3 (rechts) dem Hund (links) in Bezug auf Haarfarbe und Größe ähnlich, jedoch nicht was die Körperbehaarung oder die Kleidung betrifft.

Im Gegensatz zum Fall der Dreiecke gilt das Gesetz der Transitivität beim Vergleich anderer Objekte oft nicht. In Abb. 1.4 bestehen zum Beispiel zwischen den verschiedenen Zeichenketten paarweise Identitäten von 75%, während die erste Zeile mit der letzten keinen einzigen Buchstaben gemein hat.

Bei Betrachtung komplexerer Sachverhalte liegt somit Ähnlichkeit meist nur in einem Teil der Merkmale vor. Die Schwierigkeit besteht nun darin, zu bestimmen, welche Ähnlichkeiten bei einer bestimmten Fragestellung *wesentlich* sind und welche nicht. So ist zum Beispiel in der Philosophie die *Methode der Ähnlichkeit* eine der Methoden zur Untersuchung kausaler Zusammenhänge: Wenn zwei oder mehr Fälle einer zu untersuchenden Erscheinung nur einen Umstand gemeinsam haben, so liegt in diesem Umstand die Ursache für das Vorliegen der Erscheinung. Ist diese kausale Ähnlichkeit bekannt, so lassen sich auch Aussagen über weitere Objekte treffen, die der Ähnlichkeit genügen.

Um Ähnlichkeiten feststellen zu können, müssen Objekte verglichen werden. Bei den oben erwähnten Dreiecken ist dies keine große Herausforderung – wie erwähnt, zeigen diese nur wenige vergleichbare Merkmale. Beim Vergleich komplexerer Objekte muß hingegen eine Auswahl getroffen werden, welche Merkmale miteinander verglichen werden sollen. Um also die für uns interessanten Objekte Proteine und Kleinstrukturen vergleichen zu können, sollen hier zunächst deren Merkmale aufgezeigt werden.

H	U	N	D
H	A	N	D
H	A	N	S
H	A	S	S
F	A	S	S

Abbildung 1.4:
Intransitivität der
Ähnlichkeit bei Strings.
Die Worte sind untereinander stark ähnlich, das letzte besitzt aber keine Gemeinsamkeiten mit dem ersten.

1.4 Ähnlichkeiten zwischen Kleinstrukturen

1.4.1 Vorhandensein chemischer Gruppen

In chemischen Strukturen treten häufig wiederholende Prinzipien auf – jede Struktur lässt sich so in Bausteine zerlegen, aus denen sie zusammengesetzt ist. So besitzen zum Beispiel Steroide ein charakteristisches Ringsystem (Abb. 1.5 a). Dieser Baustein wird im Körper in Gestalt von Sexualhormonen und Kortikoiden, die den Kohlenhydrat- und Mineralstoffwechsel beeinflussen, verwendet.

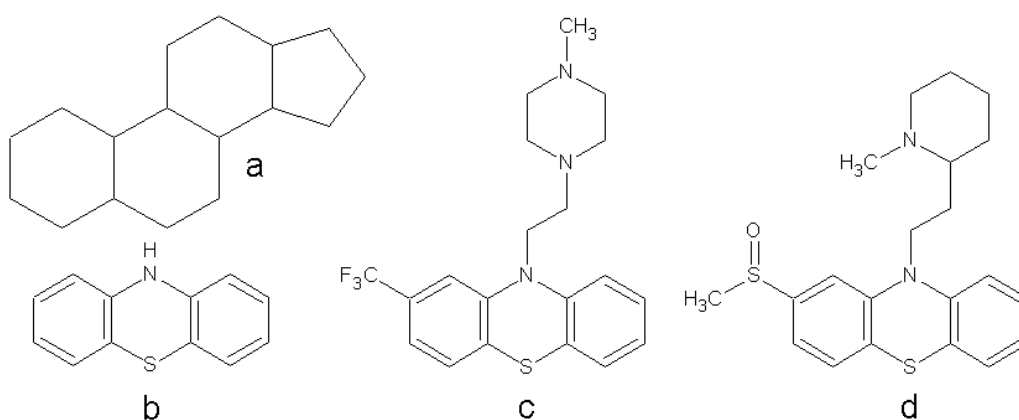


Abbildung 1.5: a) Grundgerüst der Sterole, von der sich Sexualhormone und Kortikoide ableiten. b) Grundgerüst der Phenothiazine, einer Klasse antipsychotischer Medikamente. c) Trifluoperazin, ein stark potentes Neuroleptikum mit Piperazingruppe. d) Mesoridazin, ein schwach potentes Neuroleptikum mit Piperidingruppe.

Auch bei pharmakologischen Wirkstoffen trifft man wiederkehrende Strukturelemente: So zeigen zum Beispiel die Phenothiazine, eine Gruppe antipsychotischer Wirkstoffe, eine charakteristische Dreiringstruktur (Abb. 1.5 b). Durch das Anfügen oder Entfernen von chemischen Gruppen können die Charakteristika von Medikamenten verändert werden – so kann zum Beispiel eine polare Gruppe die Löslichkeit erhöhen. Für die Wirksamkeit der Medikamente sind dabei oft zentrale Komponenten ausschlaggebend, weitere Gruppen modifizieren die Affinität oder Spezifität.

Bei Phenothiazinen führt zum Beispiel eine Piperazingruppe (z. B. im Trifluoperazin, Abb. 1.5 c) zu stark potenten Neuroleptika mit geringer sedierender Wirkung, während Komponenten mit Piperidingruppe (z. B. Mesoridazin, Abb. 1.5 d) eine starke sedierende Wirkung bei geringen extrapyramidal-motorischen Nebenwirkungen aufweisen.

Um chemische Strukturen miteinander zu vergleichen, wird das Vorhandensein der verschiedenen Gruppen in den Substanzen in einem Bitmuster, dem sogenannten *fingerprint* (Abb. 1.6 [15]), gespeichert. Dieser stellt eine Folge von 966 Bits dar, in denen gespeichert ist, ob die Struktur bestimmte Merkmale enthält oder nicht.

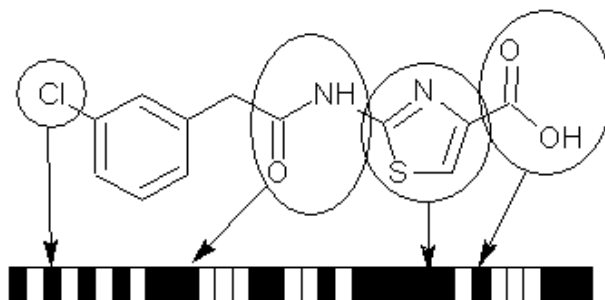


Abbildung 1.6: Fingerprint einer Kleinstruktur. Das Vorhandensein oder Fehlen bestimmter Strukturmerkmale wird in ein Bitmuster übersetzt.

Um zwei Strukturen miteinander zu vergleichen, müssen nun diese Bitmuster auf Übereinstimmungen geprüft werden. Als Ähnlichkeitsmaß hat sich der Tanimoto-Koeffizient [16] durchgesetzt:

$$T = \frac{N_{ab}}{N_a + N_b - N_{ab}}$$

mit N_{ab} = positive Bits in beiden Verbindungen, N_a = positive Bits in Verbindung a , N_b = positive Bits in Verbindung b . Durch den Koeffizienten wird die Ähnlichkeit zwischen zwei Substanzen durch eine Zahl zwischen Null (keine Übereinstimmung) und Eins (identische Strukturen) ausgedrückt, was die Analyse quantitativer Struktur-Aktivitäts-Relationen (QSAR) ermöglicht. Ein Tanimoto-Koeffizient von 0,85 kann als guter Indikator für ähnliche pharmakologische Aktivität interpretiert werden [16, 17]. Ein Vergleich mit anderen Ähnlichkeitsmaßen zeigte, daß der Tanimoto-Koeffizient am deutlichsten zwischen aktiven und inaktiven Substanzen unterscheidet [18].

1.4.2 Dreidimensionale Struktur

Die Strukturformel einer Substanz zeigt zwar deren chemische Gruppen auf, vernachlässigt aber die Anordnung der Atome im Raum. Tatsächlich beinhaltet selbst eine dreidimensionale Darstellung des Moleküls im Raum nur ein unvollständiges Modell, da die Bindungen im Molekül flexibel und teilweise

drehbar sind. Eine realitätsnahe Darstellung eines Moleküls kann durch eine Gruppe von Konformeren erfolgen, die unterschiedliche Möglichkeiten der Lage der Atome im Raum darstellen.

Jedes Konformer besitzt dabei eine charakteristische Energie, die nach der Boltzmann-Verteilung auch ein Maß dafür ist, welcher Anteil der Moleküle die jeweilige Konformation annimmt. Das Verhältnis der Auftrittswahrscheinlichkeiten in verschiedenen Zuständen hängt von deren Energiedifferenz ab:

$$\frac{N_a}{N_b} = e^{-\frac{\Delta E}{kT}}.$$

Dabei sind N_a und N_b die Häufigkeiten der Konformere a und b , ΔE ist deren Energiedifferenz, $k = 1,38 \cdot 10^{-23} \frac{J}{K}$ die Boltzmann-Konstante und T die absolute Temperatur.

Die Zahl der Konformere ist von der Anzahl der drehbaren Bindungen eines Moleküls abhängig – zum Beispiel besitzen Steroide wegen ihres starren Ringsystems nur geringe Flexibilität, während das in Abb. 1.6 gezeigte Molekül mehrere Konformationen annehmen kann.

Um Ähnlichkeiten zwischen Strukturen detektieren zu können, werden diese überlagert. Dabei wird die Struktur als Punktwolke im Raum interpretiert und die Translations- und Rotationsbewegung ermittelt, die die Strukturen möglichst gut zur Deckung bringt. Bei bekannter Zuordnung der jeweiligen Atome zueinander ist das Problem eindeutig lösbar [19]. Fehlt diese Information, sind nur Näherungslösungen möglich [20]. Letztendlich besteht somit die Schwierigkeit bei der Strukturüberlagerung im Auffinden der richtigen Zuordnung der Atome zueinander.

Dabei muß stets ein Kompromiß gefunden werden zwischen der Überlagerung möglichst vieler Atome und einer möglichst guten Deckung der überlagerten Atome miteinander. Als Qualitätskriterium für die Deckung gilt dabei der mittlere quadratische Abstand der überlagerten Atome (*root mean square deviation*, RMSD [21]). Um die Güte einer Überlagerung zu beurteilen, kann ein *score* verwendet werden, der sich aus dem Produkt des Anteils überlagelter Atome A mit der zum negativen RMSD potenzierten Eulerschen Zahl ergibt: $s = A \cdot e^{-rmsd}$ [20].

Durch 3D-Vergleiche werden so oft Ähnlichkeiten deutlich, die durch 2D-Vergleiche nicht aufgefunden werden können. Beispielhaft sind in Abb. 1.7 strukturelle Ähnlichkeiten von Neuroleptika aufgezeigt, die Nebenwirkungen erklären [20].

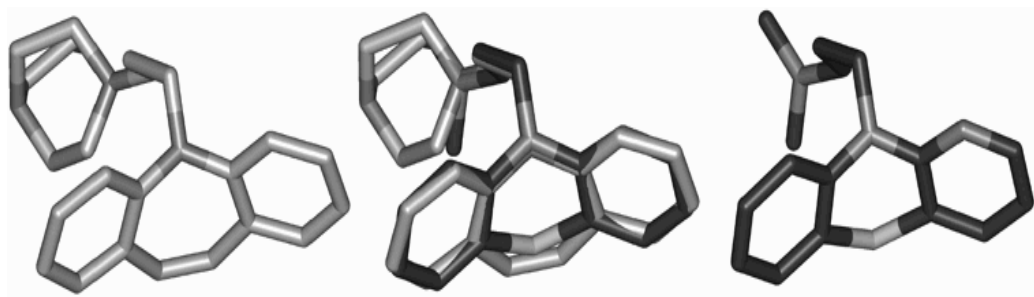


Abbildung 1.7: 3D-Ähnlichkeiten zwischen Neuroleptika. Die Ähnlichkeit des Neuroleptikums Prothipendyl (rechts) zum Antidepressivum Opipramol (links) ist durch 2D-Vergleiche nicht feststellbar – ein sechsgliedriger Ring ist durch einen siebengliedrigen Ring ersetzt. Abbildung aus [20].

1.5 Ähnlichkeiten zwischen Proteinen

1.5.1 Sequenzalignments

Wie in Abschnitt 1.1 dargelegt, ist die Zahl bekannter Proteinsequenzen überwältigend, während weitere, insbesondere strukturelle Informationen rar sind. Eine der Hauptaufgaben der Bioinformatik besteht somit darin, aus der bloßen Sequenz eines Proteins weitere, insbesondere funktionelle und strukturelle, Informationen abzuleiten. Dazu gehören die Funktion des betreffenden Proteins, seine phylogenetische Einordnung und Aussagen zur Struktur.

Der erste Schritt ist dabei stets der Vergleich einer neuen Sequenz mit bereits bekannten Proteinen, das *Alignment*. Abb. 1.8 zeigt exemplarisch ein multiples Alignment von Prionproteinen aus verschiedenen Organismen. Durch das Alignment wird nicht nur die Verwandtschaft der Proteine deutlich, es lassen sich auch Bereiche der jeweiligen Proteine einander zuordnen. Aus der Abbildung wird auch ersichtlich, daß Bereiche gleicher Sekundärstrukturen auch im Alignment einander zugeordnet werden.

Zur Erstellung eines Alignments wird eine Ähnlichkeitsmatrix zwischen Aminosäuren benötigt. Ähnliche Aminosäuren sollen möglichst einander zugeordnet werden. Da davon auszugehen ist, daß ähnliche Aminosäuren im Laufe der Evolution häufiger gegeneinander ausgetauscht werden als unähnliche, wird gern auf Austauschmatrizen zurückgegriffen. Die erste Austauschmatrix wurde 1978 von Margaret Dayhoff [23] berechnet. Aus ca. 100 Proteinen, die untereinander eine Sequenzidentität von über 85% aufwiesen, wurde die Austauschhäufigkeit der einzelnen Aminosäuren miteinander ausgezählt. Die Häufigkeit, mit der ein Aminosäurepaar bei einer angenommenen gesamten Mutationshäufigkeit von 1% ausgetauscht wird, bezeichnete sie als 1 PAM (*percent accepted mutations*).

hPrP	KK	.RP.	KP..GGWNTGGSRY	PGQGSPGGRNTP(PxGGGWGQ) ₅GGGTH.SQWN	KPSKP...KTNMK
chPrP	KK	.GKG	KPSGGGWGAGSHRQ	(PxYPxx) ₈	.GQGYNPSSGGS.YHNQ..	KPWKPP...KTNFK
tPrP	KK	.GKG	KGGGGG-NTGSNRN	(PxYPxx) ₁₀	GGQHYNPAGGDTNFKNQ..	KPWKPKDKPTNMK
xlPrP	KK	SGGG	KSKTGGWNTGSNRN	PNYP..GGYPGNT..GGSWGY	..QPYNPS.....GYN	KQWKPPKSKTNMK

hPrP	H	MAGAAAA	GAVVGGLGG	YMLGSAMS	RPIIHFGSDY	EDRYYREN	MH	RYPNQVYY	RPMDEY...SNQNN
chPrP	H	VAGAAAA	GAVVGGLGG	YAMGRVMS	GMNYHFDRPD	EYRWSEN	SA	RYPNRVYY	RDYSS...PVPQDV
tPrP	A	MAGAAAA	GAVVGGLGG	YALGSAMS	GMRMNFDRPE	ERQWNNEN	SN	RYPNQVYY	KEYNDR.SVPEGR
xlPrP	S	VA....	IGAAAGAIGG	YMLGNAV	RMSYQFNNPM	ESRYYNDY	YN	QMPNRVYR	PMYRGEEYVSEDR

hPrP	FVHDC	V	NIT	IKQHTVTTTTKG.....E	NFT	...ETDVKMMER	VVEQMCITQY	ERESQAYYQRG	S
chPrP	FVADC	F	NIT	VTEYSIGPAKKNTSEAVAAA	NQT	...EVEMENKVVTK	VIREMCVQQY	REYRLA.....	S
tPrP	FVRDC	V	NIT	VTEYKIDPNENQ.....	NVT	...QVEVRVMKQ	VIQEMCMQQY	EQYQLA.....	S
xlPrP	FVRDC	Y	NMS	VTEYIIKPAEGK.....	NNS	ELNQLDTTVKSQ	IIREMCITEY	RRG.....	S

Abbildung 1.8: Multiples Alignment von Prionproteinen aus Mensch (hPrP), Huhn (chPrP), Schildkröte (tPrP) und Frosch (xlPrP) mit Sekundärstruktur-Zuweisungen des C-terminalen Teils. Schwarz umrandet: Oligorepeats; grün umrandet: GPI-Anker; blau umrandet: Glycosylierungsstellen; gelb umrandet: konservierte Bereiche; rot umrandet: konservierte, für Stabilität wichtige Bereiche; rot: α -Helix; grün: β -sheet; magenta: 3_{10} -Helix. Die Sekundärstrukturen der alignierten Reste entsprechen einander. Abbildung nach [22].

Jedes Matrixelement ist dabei ein Maß für die Wahrscheinlichkeit, daß aus der Aminosäure x die Aminosäure y wird. Durch Multiplikation der Matrix mit sich selbst ergeben sich die angegebenen Wahrscheinlichkeiten bei einer größeren angenommenen Mutationsrate. Gebräuchlicher sind heute BLOSUM-Matrizen [24]. Sie beruhen auf lokalen Alignments aus Sequenzen mit einem bestimmten Prozentsatz an paarweiser Sequenzidentität.

Die Erstellung des Alignments entspricht somit dem Auffinden der bestmöglichen paarweisen Zuordnung von Aminosäuren der beteiligten Proteine zueinander. Der Algorithmus zum Auffinden des besten globalen Alignments zwischen zwei Sequenzen stammt von Needleman und Wunsch [25]: Für zwei Proteine der Längen n und m werden in einer $(n + 1) * (m + 1)$ -Matrix die Sequenzen der zu alignierenden Proteine waagrecht und senkrecht aufgetragen. Um *gaps* am Anfang des Alignments zu erlauben, wird in der ersten Zeile und Spalte jeweils ein Deletionssymbol eingetragen (Abb. 1.9 a).

Jedes mögliche Alignment zwischen zwei Proteinen entspricht nun einem Pfad in der zweidimensionalen Matrix. Ausgehend vom oberen linken Feld, das den Wert Null erhält, werden nun die weiteren Felder ausgefüllt. Für jedes Feld der Matrix gibt es dabei drei Möglichkeiten:

1. Man aligniert die entsprechenden Aminosäuren. Dies entspricht in der Matrix einem Schritt nach rechts unten. Der Eintrag im entsprechenden Feld ergibt sich aus der Summe des Feldes links über dem behandelten

a)	-	Q	G	T	C	
-	0	→ -1	→ -2	→ -3	→ -4	
Q	↓ -1	2	→ 1	→ 0	→ -1	
C	↓ -2	↓ 1	2	→ 1	2	b)
G	↓ -3	↓ 0	3	→ 2	→ 1	Q - G T C
T	↓ -4	↓ -1	↓ 2	5	→ 4	
C	↓ -5	↓ -2	↓ 1	↓ 4	7	Q C G T C

Abbildung 1.9: Alignment nach Needleman und Wunsch. a) komplett ausgefüllte Matrix. Pfeile zeigen an, ob jeweils eine Insertion (→), eine Deletion (↓) oder eine Alignierung (↘) vorgenommen wurde. Hervorgehoben ist der durch backpropagation aufgefundene beste Pfad; b) fertiges Alignment

Feld und dem der Paarung der Aminosäuren entsprechenden Eintrag der Substitutionsmatrix.

2. Man fügt eine Lücke, ein *gap* ein, dies entspricht in der Matrix einem Schritt nach unten. Der Wert im entsprechenden Feld ergibt sich aus dem darüber stehenden Feld abzüglich einer *gap penalty*.
3. Man fügt eine Insertion ein. Dies entspricht in der Matrix einem Schritt nach rechts. Der Eintrag im entsprechenden Feld ergibt sich aus dem links benachbarten Feld abzüglich der *gap penalty*.

Der Eintrag in jedem Feld der Matrix entspricht dem größten der drei Werte. In Abb. 1.9 ist die Matrix unter Verwendung eines *matching score* von 2, eines *mismatch score* von 0 und einer *gap penalty* von 1 ausgefüllt. Der *score* des Alignments ergibt sich dabei aus dem größten in der Matrix vorkommendem Wert (7), das Alignment selbst ergibt sich aus dem Pfad, der die größten Werte enthält und wird durch *backpropagation* gefunden. Das Alignment zwischen den beiden Sequenzen ist in Abb. 1.9 b gezeigt. Ein entsprechendes Verfahren zum lokalen Sequenzalignment geht auf Smith und Waterman zurück [26].

Statt dieser rechenaufwendigen Verfahren wird heute meist ein heuristisches Verfahren, das *basic local alignment search tool* (BLAST, [27, 28]),

verwendet. Das Verfahren beruht darauf, zunächst kurze Stücke großer Ähnlichkeit aufzufinden und daraufhin das Alignment zwischen diesen Stücken zu berechnen, und ist wesentlich schneller als der Smith-Waterman-Algorithmus.

Der Vergleich von Proteinen miteinander durch Alignments liefert also einen Hinweis auf ihre evolutionäre Verwandtschaft miteinander. Bei Proteinen mit hinreichender Sequenzähnlichkeit kann davon ausgegangen werden, daß sie eine ähnliche Faltung annehmen [29]. Dies ist durch ihre Homologie miteinander begründet, da eine zur Erfüllung der Funktion erfolgreiche Faltung im Verlauf der Evolution als konserviert angesehen werden kann.

Ähnlichkeiten in der Sequenz ohne zugrundeliegende Homologie der beteiligten Proteine müssen somit auch nicht zwangsläufig auf eine ähnliche Faltung hinauslaufen (Chamäleon-Sequenzen, s. S. 18). Ein anderes Beispiel für Ähnlichkeiten zwischen Proteinen ohne evolutionären Hintergrund stellt die häufig auftretende inverse Sequenzähnlichkeit [30] dar. Über deren strukturelle Bedeutung wird kontrovers diskutiert (s. Kapitel 5.2).

1.5.2 Sequenzmotive

Neben globalen Ähnlichkeiten zwischen Sequenzen, die durch Alignments aufgefunden werden, sind für viele Interaktionen kürzere lokale Sequenzbereiche ausschlaggebend. Als Beispiel seien hier Bereiche, die spezifisch von Binde-domänen wie PDZ und SH2 erkannt werden, genannt. Auch für posttranslationale Modifikationen wie zum Beispiel Glycosylierungen gibt es charakteristische Sequenzmotive, und lokale Strukturen wie *cis*-Prolylreste zeigen charakteristische Präferenzen in ihrer Aminosäure-Umgebung.

Die Erkennung von spezifischen Sequenzen spielt auch beim Transport eines neu synthetisierten Proteins an seinen Bestimmungsort eine Rolle. So besitzen Proteine, die in Organellen transportiert werden sollen, eine sog. *Signalsequenz*. Im Falle von Proteinen, deren Weg über das Endoplasmatische Reticulum (ER) führt, ist dies eine kurze Helix, die nach der Translokation in das Organell meist proteolytisch abgespalten wird.

Auch Mitochondrien, Chloroplasten und Peroxisomen besitzen eine Importmaschinerie, die zugehörige Proteine erkennt und in das Organell transloziert. Für Mitochondrien und Chloroplasten sowie peroxisomale Matrixproteine sind die Erkennungssequenzen gut charakterisiert, über Targetingsignale peroxisomaler Membranproteine ist bisher nur wenig bekannt. Zum Auffinden derartiger Sequenzbereiche gibt es mehrere Möglichkeiten:

Pattern-Suchen

Oft sind für die Interaktion eines Proteins mit einem Partner kurze Bereiche entscheidend. Diese Sequenzen können in verschiedenen Proteinen auftreten, die miteinander ansonsten nicht verwandt sind. Die Beschreibung eines Sequenzstückes, die für eine Interaktion erforderlich ist, bezeichnet man als Konsensussequenz oder Pattern. So ist zum Beispiel eine Dimerisierungsdomäne, der *leucine zipper*, durch die Konsensussequenz L-X₆-L-X₆-L-X₆-L gekennzeichnet, und Proteine mit der Sequenz PXXPP binden bevorzugt an SH3-Domänen. Ein Pattern enthält Aussagen darüber, welche Aminosäuren an verschiedenen Positionen des Motivs toleriert werden. Der *output* eines Patterns ist daher eine ja-nein-Entscheidung – entweder das Motiv entspricht dem Pattern oder nicht.

Hidden Markov Models (HMM) und Position Specific Scoring Matrices (PSSM)

HMM sind statistische Modelle, die einer gegebenen Sequenz eine Wahrscheinlichkeit (score) zuordnen. Im einfachsten Fall ist ein HMM ein Profil, das an mehreren Sequenzpositionen die Wahrscheinlichkeit für das Auftreten einer Aminosäure enthält. Generell bezeichnet man Matrizen, die *scores* für das Auftreten von Aminosäuren an bestimmten Positionen enthalten, als *position specific scoring matrices* (PSSM). Komplexere HMM beschreiben zusätzlich Übergangswahrscheinlichkeiten in verschiedene *states*, was zum Beispiel das Einfügen von zusätzlichen Aminosäuren ermöglicht. Im Gegensatz zu Pattern erlauben HMM das Sortieren (*ranking*) von Sequenzmotiven nach ihrer Wahrscheinlichkeit, statt nur eine ja-nein-Entscheidung zu fällen.

Neuronale Netze

Neuronalen Netze (*Artificial neural networks*, ANN) sind biologischen Schaltkreisen nachempfundene Computerprogramme. Analog zur Verknüpfung zwischen tierischen Neuronen bestehen die Netzwerke aus sogenannten Perceptrons, die Information erhalten, und mehreren Schichten (*layers*) nachgeschalteter Neuronen, die sich abhängig von ihrem Input gegenseitig beeinflussen.

Die erste Schicht, die Perceptrons, reagieren dabei auf verschiedene Eigenschaften der zu beurteilenden Sequenz (Input). Die Inputs des Perceptrons werden mittels einer Funktion aufsummiert. Der Funktionswert bestimmt dabei die Information, die an die nachgeschalteten Neuronen weitergeleitet wird, der *Output* des neuronalen Netzes ist durch den Zustand der letzten Schicht (*output layer*) bestimmt.

Neuronale Netze werden durch ein dem Lernen analogen Vorgang *trainiert*, was zur Ausbildung von Verknüpfungen zwischen den Neuronen führt. Die Methode ist sehr mächtig, bietet aber den Nachteil, daß das trainierte Netzwerk aus komplexen Verschaltungen der Neuronen besteht und seine Funktion daher oft nicht wirklich verstanden werden kann.

Support Vector Machines

Eine weitere Möglichkeit zur Vorhersage von Eigenschaften bieten die Support Vector machines. Der *Input* wird dabei als Vektor im mehrdimensionalen Raum (*input space*) interpretiert. Durch eine Rechenoperation können die Vektoren in ein anderes, meist mehrdimensionales Koordinatensystem (*feature space*) übertragen werden.

Die Idee der Support Vector Machines besteht darin, daß die transformierten Vektoren im mehrdimensionalen Raum durch eine *Hyperfläche* voneinander getrennt werden können. Zur Definition dieser Hyperfläche ist nur eine Teilmenge des Input, die *Support Vektoren*, nötig, zwischen denen die Hyperfläche aufgespannt wird. Der Trick besteht darin, daß durch die Anwendung einer Funktion (*Kernel*) direkt der Abstand eines Vektors von der Hyperfläche im *feature space* berechnet werden kann, ohne vorher eine Transformation in dieses Koordinatensystem vornehmen zu müssen. Als *output* wird letztlich ausgegeben, auf welcher Seite der Hyperfläche sich der zu einer gegebenen Sequenz zugehörige Vektor befindet, es erfolgt also eine ja- nein-Klassifikation.

1.5.3 Strukturvergleiche

Bei der ersten Betrachtung mag eine Proteinstruktur wie eine zufällige Anordnung der Aminosäurekette im Raum wirken. Die Aufklärung der Struktur des Myoglobins im Jahre 1958 [31] war daher ein Schock für die Wissenschaftler, da sich die Hoffnung nach einfachen Prinzipien der Struktur, wie in der DNA, nicht erfüllte. John Kendrew, der die Struktur aufklärte, kommentierte seine Entdeckung mit den Worten[32]:

„Perhaps the most remarkable features of the molecule are its complexity and its lack of symmetry. The arrangement seems to be almost totally lacking the kind of regularities which one instinctively anticipates, and it is more complicated than has been predicted by any theory of protein structure.“

Tatsächlich zeigen Proteinstrukturen zwar einen weit komplexeren Aufbau als die Doppelhelix der DNA, bei näherer Betrachtung zeigen sich jedoch

wiederkehrende Prinzipien. So besteht zum Beispiel das Innere von Proteinen meist aus hydrophoben Aminosäuren, während die Oberfläche eher polar ist. Man geht davon aus, daß die Wechselwirkungen der hydrophoben Aminosäuren miteinander bei der Proteinfaltung eine wichtige Rolle spielen – ein Kontakt zwischen hydrophoben Aminosäuren ist energetisch günstiger als die Exposition der gleichen Seitenketten zum Wasser.

Um allerdings die hydrophoben Seitenketten eines Proteins in dessen Innerem zu konzentrieren, muß auch das *backbone* des Proteins sich im hydrophoben Zentrum befinden. Die Hauptkette ist allerdings polar und daher hydrophil – auf jede Aminosäure kommt ein Wasserstoffbrücken-Donor (NH) und ein Akzeptor (CO). Um diese Gruppen in einer hydrophoben Umgebung zu stabilisieren, müssen die Wasserstoffbrücken-Donoren und -Akzeptoren miteinander in Kontakt treten. Dieses Problem wird durch die Ausbildung von *Sekundärstrukturen* gelöst.

Schon 1951 sagte Linus Pauling aufgrund der Notwendigkeit der Ausbildung von Wasserstoffbrücken zwischen CO- und NH-Gruppen zwei theoretische periodische Strukturen voraus – die α -Helix [33]¹ und das β -Faltblatt [34]. In der rechtsgängigen α -Helix (Abb. 1.10) bildet jeweils die CO-Gruppe der Aminosäure an Position i eine Wasserstoffbrücke zur NH-Gruppe der Aminosäure an Position $i + 4$ aus. Die Ganghöhe der Helix entspricht 5,4 Å, eine Windung entspricht ca. 3,6 Aminosäuren, die Seitenketten der Aminosäuren zeigen nach außen.

α -Helices in Proteinen sind oft amphipatisch, d. h. sie besitzen eine hydrophobe und eine hydrophile Seite. Auf diese Weise kann eine Seite der Helix an der Außenseite des Proteins liegen und die andere zum hydrophoben Kern hin orientiert sein. Andererseits kann auch durch entsprechende Wahl der Seitenketten eine komplett hydrophobe Oberfläche geschaffen werden, die es der Helix ermöglicht, die Membran zu durchspannen und so das Protein durch eine *Transmembranhelix* zu verankern.

Neben der α -Helix treten zu kleinen Teilen auch Helices auf, bei denen die Wasserstoffbrückenbindung zum Rest $i + 3$ (3_{10} -Helix) oder $i + 5$ (π -Helix) ausgebildet wird. Erstere ist steiler und schmaler, letztere breiter als die α -Helix.

Die zweite von Linus Pauling postulierte Struktur ist das β -Faltblatt (Abb. 1.11). Wasserstoffbrückenbindungen werden dabei zwischen parallel oder antiparallel verlaufenden Peptid-Hauptketten gebildet, die Seitenketten befinden sich ober- und unterhalb des Faltblattes. Auf diese Weise gelingt es, die hydrophile Hauptkette im Inneren der Struktur abzuschirmen.

¹Interessanterweise ist die ursprünglich von Pauling vorgeschlagene Struktur eine linksgängige Helix – er war von D-Aminosäuren als Proteinbausteinen ausgegangen.

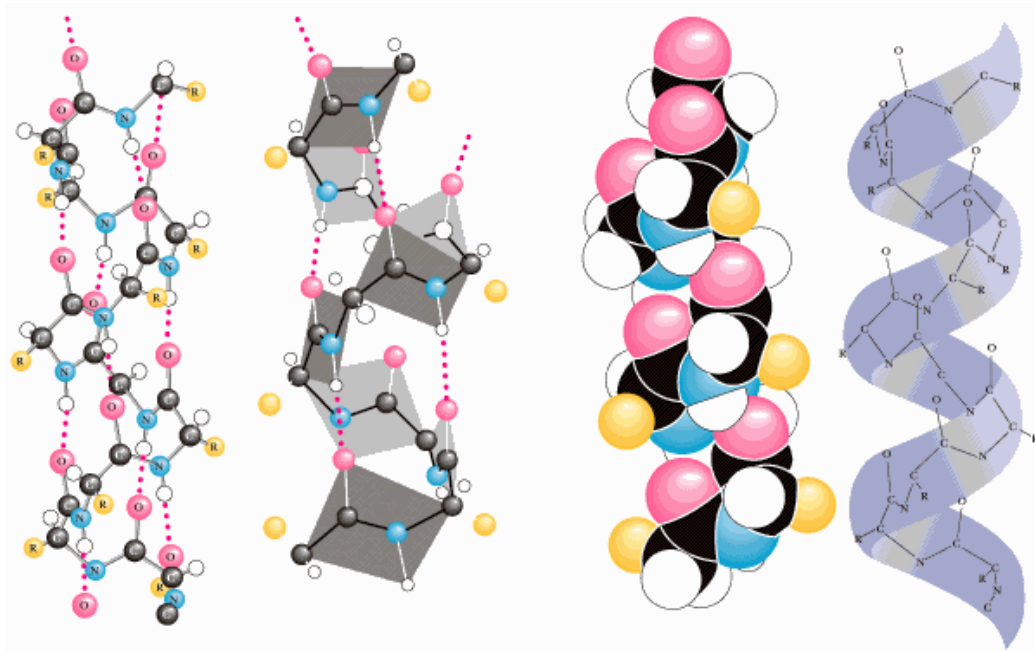


Abbildung 1.10: Struktur der α -Helix. Wasserstoffbrücken-Donoren und -Akzeptoren des Peptidrückgrats sind einander zugewandt. Abbildung aus [35].

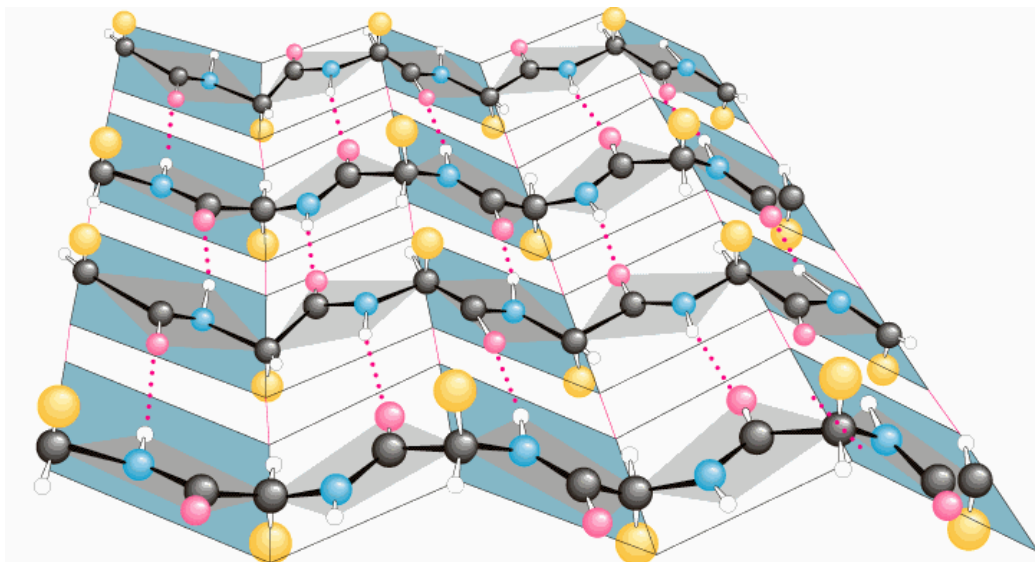


Abbildung 1.11: Struktur der β -Faltblattes. Gezeigt ist ein antiparalleles Faltblatt. Wasserstoffbrückenbindungen bilden sich zwischen verschiedenen Teilen des Peptidrückgrats aus. Abbildung aus [35].

α -Helices und β -Sheets sind stets wiederkehrende Motive in Proteinstrukturen. Anhand des Anteils dieser Sekundärstrukturen lassen sich Proteine in Gruppen teilen. So gibt es Proteinfamilien, die nur oder zum größten Teil aus α -Helices bestehen (z. B. Globine), Proteine mit größtenteils β -Sheets (z. B. Immunglobuline) und Mischformen mit hauptsächlich parallelen (z. B. TIM-Barrel) oder hauptsächlich antiparallelen (z. B. Lysozym) Faltblättern.

Die Ausbildung von Sekundärstrukturen hat unmittelbaren Einfluß auf die lokale Faltung des Peptidrückgrats. Um diese zu beschreiben, bedient man sich der Torsionswinkel ϕ , ψ und ω (Abb. 1.12).

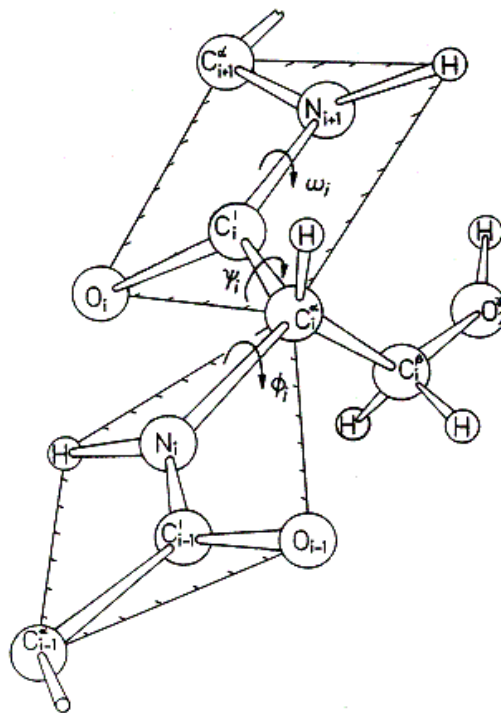


Abbildung 1.12: Torsionswinkel ϕ , ψ und ω in der Hauptkette von Proteinen. Die Winkel stellen jeweils die Drehung um eine Bindung dar.

Schon in frühen Studien konnte gezeigt werden, daß die Torsionswinkel von Proteinen nicht alle möglichen Werte annehmen können [36, 37]. Trägt man die Winkel ϕ und ψ gegeneinander auf, so ergibt sich der sogenannte Ramachandran-Plot (Abb. 1.13). Bestimmte Regionen des Plots sind charakteristisch für α -Helices bzw. β -Sheets, andere Regionen sind wegen sterischer Behinderung der Seitenketten untereinander nicht besetzt.

Neben den im Ramachandran-Plot aufgetragenen Winkeln ϕ und ψ gelten auch für den Winkel ω Restriktionen. Wegen des partiellen Doppelbin-

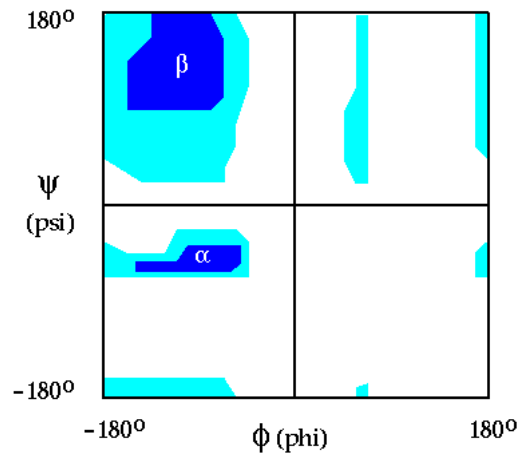


Abbildung 1.13: Ramachandran-Plot der Torsionswinkel ϕ und ψ in Proteinen. Aufgrund der Ausbildung von Sekundärstrukturen sind entsprechende Kombinationen von ϕ und ψ -Winkeln vorherrschend (blau), andere Regionen sind verboten.

dungscharakters der Peptidbindung treten hier nur Winkel um 0 (*cis*) und Winkel um 180 (*trans*) auf. Während der größte Teil der Peptidbindungen die *trans*-Konformation einnimmt, treten vor Prolylresten zu ca. 5% *cis*-Peptidylbindungen auf [38]. Offensichtlich spielt die lokale Sequenzumgebung für die Konformation der Prolylbindung eine Rolle [39].

Lokale Konformationspräferenzen scheinen generell häufig durch die Sequenz codiert zu werden [40, 41], und helikale Bereiche von Proteinen zeigen oft auch als Peptide helikale Konformation [42]. Minor und Kim konnten allerdings zeigen, daß eine elf Aminosäuren lange Sequenz abhängig von ihrer Sequenzumgebung im *host*-Protein α -helikale oder β -Faltblatt-Struktur annehmen kann (*Chamäleon-Sequenz* [43]). Offensichtlich bestimmen daher auch *long-range*-Effekte die Ausbildung von Sekundärstrukturen.

Auch eine starke Sequenzähnlichkeit zwischen Proteinen muß nicht zwangsläufig zu einer ähnlichen Struktur führen: aufgrund einer Wette synthetisierter Dalal und Mitarbeiter ein Protein mit 50% Sequenzidentität zu einem β -Sheet-Protein, das α -helikale Struktur aufweist [44]. Eine Suche in der PDB [45] führte weitere Chamäleon-Sequenzen mit einer Länge von bis zu sieben Aminosäuren zutage. Eine bloße lokale Sequenzähnlichkeit läßt demnach noch nicht auf die Faltung schließen.

Am deutlichsten wird die Möglichkeit von unterschiedlichen Faltungstypen bei gleicher Sequenz bei den Amyloidosen, einer Gruppe von Krankheiten, die auf Fehlfaltung von Proteinen beruhen. So kommt zum Beispiel das Prionprotein in einer nativen (PrP^C) und einer pathologischen Form (PrP^{Sc})

vor. Die Umwandlung von PrP^C in PrP^{Sc} wird dabei durch schon vorhandenes PrP^{Sc} katalysiert (s. Kapitel 2).

Daher ist der Vergleich der Raumstrukturen von Proteinen oft aufschlußreicher als der bloße Sequenzvergleich. Insbesondere bei paarweisen Sequenzidentitäten unter 35% (sog. *twilight zone* von 20–35% [46]) kann die Verwandtschaft zwischen Proteinen nicht mehr eindeutig durch Sequenzalignments belegt werden. Nur ein Vergleich der Proteinstrukturen läßt dann Rückschlüsse auf eine Homologie der Proteine zu. Ein Beispiel für eine stark diverse Proteinfamilie sind proteasomale Untereinheiten mit paarweisen Sequenzidentitäten von teilweise unter 10% [47]. Selbst Proteine ohne detektierbare Sequenzidentität können ähnliche Strukturen annehmen [48].

Zur Überlagerung von Proteinen kann der gleiche Algorithmus wie bei der Überlagerung von Kleinstrukturen verwendet werden [20]. Zum Auffinden eines kleinen Teils, wie zum Beispiel einer Nucleotid-Bindestelle, in einer Proteinstruktur wurde der *Needle-Haystack*-Algorithmus entwickelt [49].

Da allerdings Proteine eine Kette von Aminosäuren darstellen, ist es oft hilfreich, bei einer Überlagerung die Reihenfolge der Aminosäuren in den zu vergleichenden Proteinen zu beachten. Dies wird beim Verfahren der *combinatorial extension* (CE [50]) berücksichtigt. Dabei werden zunächst Fragmente der Proteine mit strukturellen Ähnlichkeiten (*aligned fragment pairs*, AFP) gesucht, welche dann zu einem strukturellen Alignment verknüpft werden. Eine weitere Herangehensweise besteht im Vergleich von Distanzmatrizen (*distance matrix alignment*, DALI [51]) zwischen C_α -Atomen von Proteinen.

1.6 Zielsetzung der Arbeit

In dieser Arbeit sollen mehrere Aspekte der Ähnlichkeit zwischen Proteinen und Kleinstrukturen untersucht werden, insbesondere unter Beachtung der Fragestellung, welche Schlußfolgerungen aus diesen Ähnlichkeiten gezogen werden können.

Im ersten Abschnitt stehen Ähnlichkeiten zwischen Kleinstrukturen im Vordergrund. Ausgehend von einer publizierten Liste von Inhibitoren der Aggregation des Prionproteins soll eine Datenbank pharmazeutischer Wirkstoffe nach ähnlichen Substanzen durchsucht werden. Aufgrund von 2D- und 3D-Ähnlichkeiten sollen neue Inhibitoren vorgeschlagen werden, die als Leitstrukturen zur Entwicklung von Therapeutika gegen Prionenkrankheiten dienen können. Es soll untersucht werden, welche strukturellen Merkmale bekannte Inhibitoren gemein haben, und wie aus dem Vorhandensein dieser Merkmale auf eine Wirksamkeit gegenüber Prionenkrankheiten geschlossen werden kann.

Im zweiten Teil sollen Ähnlichkeiten zwischen Sequenzmotiven herausgearbeitet werden, die eine Interaktion von peroxisomalen Membranproteinen mit deren Importrezeptor Pex19 vermitteln. Während andere Targetingsignale für Proteine bereits gut charakterisiert ist, weiß man noch sehr wenig über Targetsignale peroxisomaler Membranproteine. Auf Grundlage von Daten aus Peptid-Spot-Experimenten soll die Bindestelle charakterisiert und eine Vorhersagematrix zum Auffinden von Bindestellen für Pex19 erstellt werden. Welche Aminosäuren sind für das Targeting wichtig? Wie ähnlich sind sich die Targetsignale unterschiedlicher Proteine?

Der dritte Teil der Arbeit untersucht den Einfluß lokaler Sequenzähnlichkeit auf die Konformation von Peptidylbindungen. Es soll herausgearbeitet werden, inwieweit die lokale Sequenzumgebung ein Indikator für die Konformation der Peptidbindung von Prolylresten darstellt, und welche Effekte Aminosäuren in der Nachbarschaft des Prolylrestes bewirken. Informationen über den Einfluß der lokalen Sequenzumgebung auf die Proteinfaltung können helfen, die für die Faltung wichtigen Wechselwirkungen zu verstehen und aus der Aminosäuresequenz Aussagen über die Struktur eines Proteins zu treffen.

Der letzte Teil der Arbeit beschäftigt sich mit inverser Sequenzähnlichkeit zwischen Proteinen und beleuchtet die Frage, ob diese ein Indikator für strukturelle Ähnlichkeit der beteiligten Proteine ist. Im Falle einer strukturellen Bedeutung inverser Sequenzähnlichkeit könnte die Lernmenge beim Vergleich von Proteinsequenzen verdoppelt werden. Es soll untersucht werden, inwieweit Proteine mit inverser Sequenzähnlichkeit ähnliche Faltungen annehmen. Inverse Sequenzähnlichkeit ist dabei auch ein Modell für Sequenzähnlichkeiten ohne zugrundeliegende Homologie, über deren Einfluß auf die Struktur wenig bekannt ist.

Kapitel 2

Kleinstrukturen: Inhibition der Prionenaggregation

2.1 Fehlfaltung von Proteinen

Der native Zustand eines Proteins entspricht entgegen der Hypothese von Anfinsen [52] nicht immer dem thermodynamisch stabilsten Zustand. Fehler bei der Proteinfaltung führen teilweise zu nicht funktionellen Proteinen, im ungünstigen Fall bilden die fehlgefalteten Proteine Aggregate, die sich in der Zelle anhäufen und toxisch sind.

Eine Gruppe von Krankheiten, die auf Fehlfaltung von Proteinen beruhen, sind die Amyloidosen (zur Übersicht s. [53]). Gemeinsames Merkmal dieser Erkrankungen ist, daß sich ein Protein durch Fehlfaltung in unlöslichen Fibrillen ablagert, die reich an β -Sheet-Strukturen (s.S. 15) sind, sich durch Kongorot anfärben lassen und dabei eine charakteristische Doppelbrechung zeigen.

Die bekanntesten Amyloidosen sind die Alzheimer-Krankheit, bei der sich ein Fragment des Alzheimer-Precursor-Proteins (APP) ablagert, und die Transmissiblen Spongiformen Encephalopathien (TSE), bei denen sich eine fehlgefaltete Form des Prionproteins bildet, was zu Neurodegeneration führt. Zu dieser Gruppe von Erkrankungen gehören BSE, die Creutzfeld-Jakob-Krankheit (CJD), das Gerstmann-Sträussler-Scheinker-Syndrom (GSS), die Fatale Familiäre Insomnie (FFI), Kuru und Formen bei Schafen und Ziegen (Traberkrankheit, *Scrapie*), Nerzen, Hirschen und Katzen. Auch die Parkinson-Krankheit, Diabetes Typ II und das Down-Syndrom gehören neben weiteren Krankheiten zu den Amyloidosen.

In diesem Abschnitt sollen die Prionerkrankungen (TSE) näher erläutert werden. Die Kennzeichen sind Vakuolisierung im Gehirn, Astrogliose, neurale

Apoptose und die Anhäufung von proteaseresistentem Prionprotein (PrP^{res}) im zentralen Nervensystem. Wegen der ungewöhnlich langen Inkubationszeit wurde zunächst ein langsamer Virus für den Verursacher der Krankheit gehalten [54]. Alper und Mitarbeiter konnten jedoch bereits in den 60er Jahren zeigen, daß der Erreger extrem widerstandsfähig gegen UV und ionisierende Strahlung ist – Verfahren, die Nucleinsäuren beschädigen und daher gegen Viren wirksam sind [55].

Nachdem bereits 1967 von Griffith die Möglichkeit in Betracht gezogen wurde, daß der Scrapie-Erreger ein Protein sein könnte, das sich selbst repliziert [56], wurde diese Hypothese 1982 von Stanley Prusiner aufgegriffen, der das Agens als *proteinaceous infectious particle* (Prion) bezeichnete [57]. Diese sogenannte *protein only hypothesis* setzte sich mehr und mehr durch.

Nach der Hypothese führt der Kontakt der zellulären Form des Prionproteins (PrP^{C}) mit der fehlgefalteten, infektiösen Form (PrP^{Sc}) zu einer Umfaltung von PrP^{C} zu PrP^{Sc} und damit zur Propagation der Erkrankung. Dieses Modell wird durch mehrere Evidenzen gestützt. So sind zum Beispiel erbliche Formen von TSE stets mit einer Mutation des Prionproteins verknüpft [58], und die Expression eines PrP -Gens mit einer GSS-assoziierten Punktmutation in Mäusen führte zu neurologischer Dysfunktion und spongiformer Neurodegeneration [59]. Mäuse, die das PrP -Gen nicht exprimieren, sind hingegen resistent gegen Scrapie [60]. Schließlich konnte die katalytische Wirkung von PrP^{Sc} bei der Umfaltung von PrP^{C} auch *in vitro* gezeigt werden [61, 62]. In jüngster Zeit gelang auch die Erzeugung infektiösem Materials *in vitro* [63].

Die strukturellen Änderungen bei der Umfaltung von PrP sind noch unverstanden. Durch CD-Spektroskopie konnte jedoch gezeigt werden, daß der Anteil an β -Faltblättern während der Konversion des Prionproteins von der nativen Form (PrP^{C}) in die fehlgefaltete Form (PrP^{Sc}) zunimmt, während der Gehalt an α -Helices sinkt [64]. Die Struktur des nativen Prionproteins ist inzwischen bekannt, vom fehlgefalteten Protein existieren nur theoretische Modelle (Abb. 2.1).

2.2 Inhibitoren der Bildung von PrP^{Sc}

Zur Auffindung von Leitstrukturen zur Therapie der Prionerkrankungen wird derzeit nach Substanzen gesucht, die die Umfaltung von PrP in die pathologische Form verhindern und/oder bestehende Aggregate auflösen.

Daß eine Rückfaltung von PrP^{Sc} in eine lösliche, PrP^{C} -ähnliche Form möglich ist, konnte durch Experimente mit Guanidinhydrochlorid gezeigt werden [66]. Eine Inhibition der Fehlfaltung *in vitro* gelang auch durch syn-

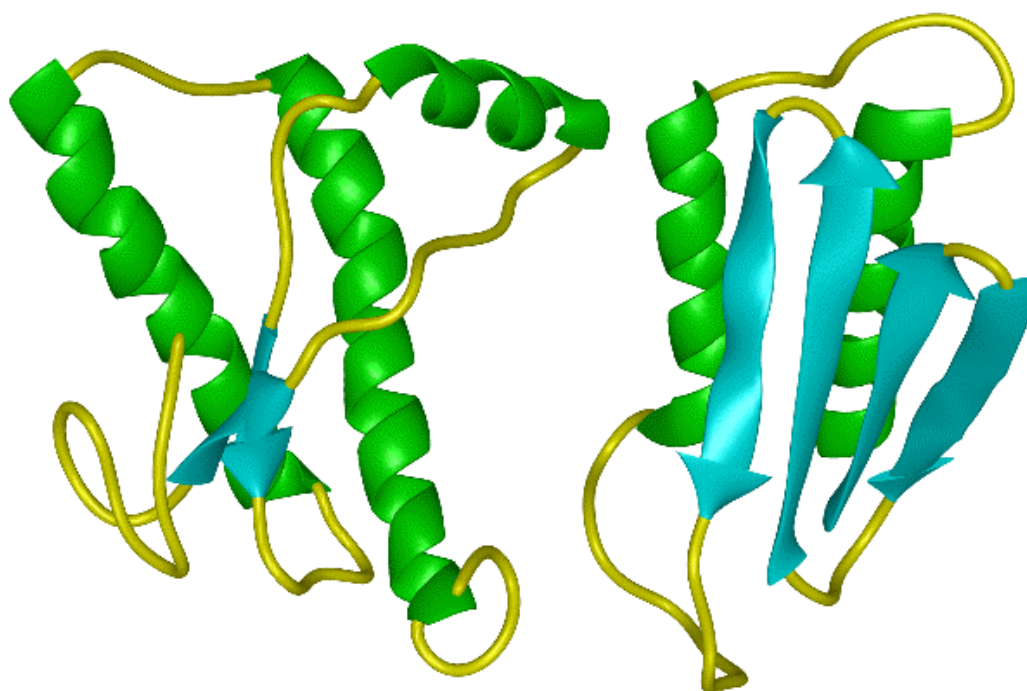


Abbildung 2.1: Kristallstruktur des nativen Prionproteins (links) und ein Modell der fehlgefalteten Form (rechts) nach [65]. Bei der Umfaltung erhöht sich der Anteil an β -Faltblättern.

thetische Peptidfragmente des Prionproteins [67, 68]. Auch Antikörper gegen PrP inhibieren die Propagation von PrP^{Sc} *in vitro* [69], in Zellkultur [70, 71] und *in vivo* [72].

Als Leitstrukturen zur Suche von Pharmaka sind hingegen weder das denaturierende Agens Guanidinhydrochlorid noch Peptide oder Antikörper geeignet. Hier interessieren im Gegenteil kleinere Moleküle. Als Inhibitoren wurden unter anderem Porphyrine und Phthalocyanine [73], verzweigtkettige Polyamine [74] und Pentosan-Polysulfat [75] diskutiert. Porphyrine und Phthalocyanine [76] sowie Pentosan-Polysulfat [77] wurden auch in Scrapie-infizierten Tieren getestet und verlängerten deren Lebenszeit deutlich.

Da allerdings die oben genannten Substanzen für eine pharmakologische Verwertung wegen ihrer Größe nicht in Betracht gezogen werden können, da sie die Blut-Hirn-Schranke nicht überwinden, sind Substanzen mit noch kleinerem Molekulargewicht von größerem Interesse. Hier sind als Beispiel der Azofarbstoff Kongorot und dessen Derivate [78, 79], Curcumin [80] sowie Substanzen aus der Gruppe der Acridine und Phenothiazine [81, 82], zu denen unter anderem das Malariamedikament Quinacrin und das Neuroleptikum

Chlorpromazin gehören, zu nennen.

Kongorot zeigte im Tierversuch nur bei direkter Injektion ins Gehirn zum Zeitpunkt der Infektion Erfolg [83]. Modifikationen der Seitenkette [84] führten allerdings zu Derivaten mit besserer Hirngängigkeit [85]. Quinacrin zeigte im Tiermodell keinen therapeutischen Effekt [86], bei Patienten mit Creutzfeldt-Jakob-Krankheit wurde allerdings eine Besserung festgestellt [87, 88]. Auch Suramin, das den intrazellulären Proteintransport beeinflusst und so zu einer Aggregation von PrP in Golgi-Vesikeln und nachfolgendem Abbau in Lysosomen führt, verhindert die Akkumulation von PrP^{Sc} in ScN2a-Zellen und verlängert die Inkubationszeit der Erkrankung in Mäusen [89].

2.3 Zielsetzung

Ausgehend von bekannten Inhibitoren der PrP^{Sc}-Bildung sollen durch *screening* einer Substanzdatenbank neue potentielle Leitstrukturen zur Therapie von Transmissiblen Spongiformen Encephalopathien vorgeschlagen werden.

2.4 Ausgangspunkte für neue Leitstrukturen

Als Grundlage für die Untersuchung dient eine Arbeit von Kocisko und Mitarbeitern [90], die systematisch eine Kollektion von 2000 Substanzen auf die Inhibition der PrP^{Sc}-Bildung in ScN2a-Zellen untersuchte. Die Substanzbibliothek der Firma Microsource Discovery (MD) enthält pharmakologische Wirkstoffe und Naturstoffe.

Als Testsystem wurden ScN2a-Zellen gewählt, die in 96-well-Platten kultiviert, mit den zu testenden Substanzen inkubiert und anschließend durch einen *dot blot* untersucht wurden. Dazu wurden die Zellen lysiert, mit Proteinase K behandelt und das Lysat auf eine PVDF-Membran übertragen. Nach Behandlung mit Guanidiniumthiocyanat (GdnSCN) und mehreren Waschschritten wurde das verbleibende PrP^{Sc} durch Antikörper detektiert. Zusätzlich wurden auf jeder Membran unbehandelte Zellen (Negativkontrolle) und mit Curcumin behandelte Zellen (Positivkontrolle) aufgetragen.

Zunächst wurden die Substanzen in einer Konzentration von 10 μ M gegen Zellen mit dem RML-Stamm [91] getestet. Ca. 70% der Substanzen zeigte weniger als 50% Inhibition der PrP^{Sc}-Bildung, 20% waren cytotoxisch und wurden nochmals in einer Konzentration von 1 μ M getestet, und 246 Substanzen inhibierten die PrP^{Sc}-Formation zu über 50%. Diese wurden daraufhin gegen Zellen getestet, die mit dem Stamm 22L transfiziert waren. 40 der 246 Substanzen zeigten auch hier eine Inhibierung von über 50% und wurden in

einer Konzentration von $1\text{ }\mu\text{M}$ gegen beide Stämme getestet.

Fünf dieser Substanzen sowie 12 der in höherer Konzentration cytotoxischen Wirkstoffe zeigten einen IC_{50} von $\leq 1\text{ }\mu\text{M}$ gegen beide Stämme. Die 17 Substanzen sind in Abb. 2.2 gezeigt. Diese Substanzen sollen als Ausgang für eine Suche in der Kleinstrukturdatenbank dienen.

2.5 Kleinstruktur-Datenbanken

Kernaufgabe der pharmakologischen Forschung ist die Suche nach neuen Strukturen, die größere Spezifität und/oder geringere Nebenwirkungen als bekannte Wirkstoffe zeigen. Eine weit verbreitete Methode ist dabei das *high throughput screening* (HTS) von Substanzdatenbanken. Dies ist allerdings zeit- und kostenintensiv, und zudem ist eine hochdurchsatzfähige Testmethode erforderlich. Daher gewinnt das *in silico* Design neuer Wirkstoffe immer mehr an Bedeutung. Das Prinzip des Verfahrens beruht auf der Erfahrung, daß strukturell ähnliche Substanzen oft auch ähnliche Wirkungsweisen zeigen.

Während vor einigen Jahren Datenbanken chemischer Strukturen nur großen Pharmaunternehmen zugänglich waren, ist die öffentliche Verfügbarkeit derartiger Tools in den vergangenen Jahren kontinuierlich gestiegen [92]. So beinhaltet zum Beispiel *Chemicval Abstracts Search* (CAS, www.cas.org) Informationen über 25 Millionen Substanzen, von denen 8 Millionen kommerziell verfügbar sind.

Zur Struktursuche wurde die in der Arbeitsgruppe Preißner erstellte Super-Drug-Datenbank [93] benutzt. Die Datenbank enthält ca. 2400 Strukturen und 100 000 Konformere pharmakologischer Wirkstoffe, die von der WHO als essentiell eingestuft werden [94]. Zusätzlich sind die medizinischen Anwendungsbereiche als sog. ATC-Code (*Anatomic-therapeutic-chemical classification of drugs* [95]) hinterlegt. Die Klassifizierung beschreibt das anatomische System, auf das ein Medikament wirkt (z. B. Nervensystem), die therapeutische Indikation (z. B. Analgetika), die chemische Klassifizierung (z. B. Opioide) und die genaue Substanzklasse (z. B. Benzomorphan-Derivate) eines Wirkstoffs.

2.6 2D-Struktursuche

Zunächst wurde mit den von Kocisko [90] identifizierten Faltungsinhibitoren eine chemische Ähnlichkeitssuche in der Datenbank durchgeführt (s. S. 7). Sechs der von Kocisko identifizierten Substanzen sind Naturstoffe, elf stellen

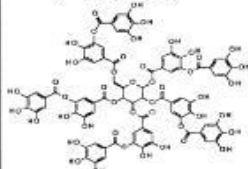
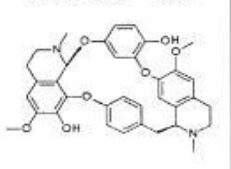
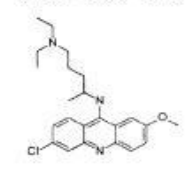
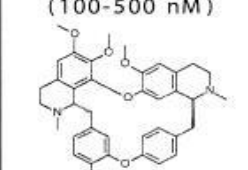
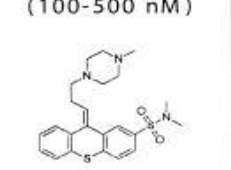
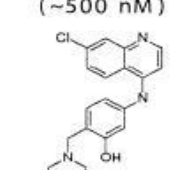
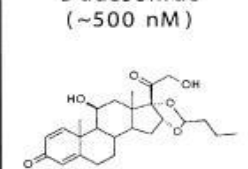
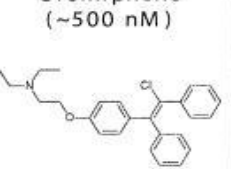
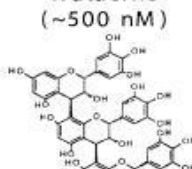

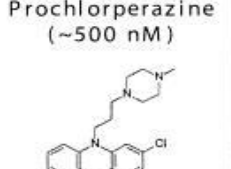
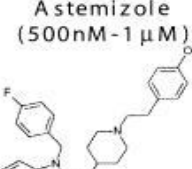
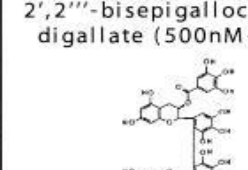

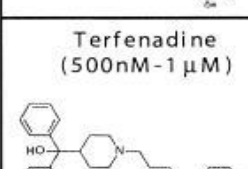
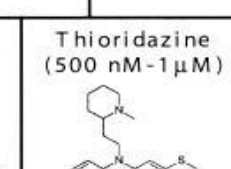
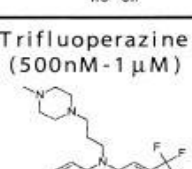
Tannic acid (~100 nM) 	Bebeerine (curine) (100-500 nM) 	Quinacrine (100-500 nM) 
Tetrandrine (100-500 nM) 	Thiothixene (100-500 nM) 	Amodiaquine (~500 nM) 
Budesonide (~500 nM) 	Clomiphene (~500 nM) 	Katacine (~500 nM) 
Lovastatin (~500 nM) 	Prochlorperazine (~500 nM) 	Astemizole (500nM-1 µM) 
2',2'''-bisepigallocatechin digallate (500nM-1 µM) 	Chrysanthellin A (500nm-1 µM) 	
Terfenadine (500nM-1 µM) 	Thioridazine (500 nM-1 µM) 	Trifluoperazine (500nM-1 µM) 

Abbildung 2.2: Inhibitoren der Fehlfaltung des Prionproteins aus [90]. 12 der 17 Substanzen mit IC_{50} eignen sich als Targets für eine Kleinstruktursuche.

Wirkstoffe aus verschiedenen Indikationsgruppen dar: Neben den schon als Fehlfaltungsinhibitoren bekannten Phenothiazinen [81] und dem Malariamedikament Quinacrin [82] finden sich auch ein Steroid (Budesonid), Lovastatin und das Antihistaminikum Astemizol. Die Dreiringstruktur des Quinacrins ähnelt der Struktur der Phenothiazine, und Teile der Struktur sind auch in dem Malariamedikament Amodiaquin konserviert. Sechs der elf Substanzen (Quinacrin, vier Phenothiazine und Amodiaquin) fallen somit in eine Gruppe.

Da die sechs von Kocisko identifizierten Naturstoffe größere Strukturen mit einem Molekulargewicht außerhalb der pharmakologisch interessanten Grenzen darstellen, wurden diese in der Untersuchung nicht berücksichtigt. Mit den von Kocisko gefundenen Arzneimittelwirkstoffen wurde daraufhin eine 2D-Suche in der SuperDrug-Datenbank (SD) und einer neuen Datenbank mit den 2000 von Kocisko untersuchten Substanzen durchgeführt. Tab. 2.1 zeigt die besten 20 Treffer jeder Suche.

Es fällt auf, daß einige Substanzen wie Budesonid zu vielen Treffern mit guter Bewertung führen, während andere, wie z. B. Clomiphen, nur wenige strukturell ähnliche Treffer liefern. Grund dafür ist die große pharmakologische Bandbreite an Steroiden, zu denen Budesonid gehört. Andererseits sind Treffer aus dem MicroSource-Set mit guter Bewertung oft Substanzen, die in Kociskos Assay Wirkung zeigen: Eine Suche mit Amodiaquin liefert beispielsweise Quinacrin, Chloroquin und Hydroxychloroquin, und Prochlorperazin liefert als Suchergebnisse weitere, schon von Kocisko als aktiv charakterisierte Phenothiazine. Sechs der elf von Kocisko beschriebenen Substanzen (Amodiaquin, Prochlorperazin, Quinacrin, Thioridazin und Trifluoperazin) sind unter den besten 20 Treffern von Suchen mit anderen Wirkstoffen aus der Gruppe.

Während aber zum Beispiel die Klasse der Phenothiazine viele aktive Substanzen beinhaltet, sind viele Steroide trotz ihrer hohen Ähnlichkeit zum wirksamen Budesonid nicht aktiv (Tab. 2.1). Eine bloße 2D-Ähnlichkeit nach Tanimoto ist also noch kein Garant für die Wirkung einer Substanz, die Bedeutung der Ähnlichkeit hängt im Gegenteil stark von der untersuchten Substanzklasse ab.

Zur Bewertung der Treffer ist somit ein struktureller Vergleich mit anderen aktiven Substanzen hilfreich – auf der anderen Seite liefert der Vergleich mit inaktiven Substanzen Aufschluß darüber, welche strukturellen Modifikationen die Aktivität einer Leitsubstanz verringern. So sind zum, Beispiel polare Endgruppen bei Phenothiazinen scheinbar mit einem Aktivitätsverlust korreliert.

Tabelle 2.1: 2D-Suchen mit den von Kocisko identifizierten Substanzen in der SuperDrug-Datenbank (SD) und den Substanzen von MicroSource (MS). *Blau: inaktiv in Kociskos Assay; Fett: IC₅₀ von $\leq 10 \mu M$ in Kociskos Assay; Fett grün: IC₅₀ von $\leq 1 \mu M$ in Kociskos Assay; Fett rot: als neue Leitstruktur vorgeschlagen*

Amodiaquine SD		Astemizole SD		Budesonide SD	
Amodiaquine	100.00	Astemizole	100.00	Budesonide	100.00
Quinacrine	67.53	Pimozide	77.30	Desonide	93.68
Chloroquine	63.24	Mizolastine	76.60	Flurandrenolide	91.67
Hydroxychloroquine	61.24	Benperidol	75.19	Flunisolide	91.31
Glafenine	59.25	Droperidol	72.98	Ciclesonide	89.71
Primaquine	59.14	Domperidone	72.94	Fluocinolone	89.43
Tretoquinol	57.49	Emedastine	71.93	Flucoronide	87.95
Gefitinib	56.73	Sertindole	71.10	Halcinonide	87.19
Chlorproethazine	55.64	Abacavir	69.86	Fluocinonide	86.06
Dequalinium	54.98	Ziprasidone	68.73	Formocortal	85.17
Eupaverin	54.37	Ketanserin	68.06	Cymarine	85.13
Ethacridine	54.19	Bezitramide	67.97	Amcinonide	84.50
Clebopride	54.14	Telmisartan	67.61	Hydrocortisone	83.95
Nomifensine	53.57	Proglumetacin	66.77	Peruvoside	83.76
Papaverine	53.39	Clozapine	66.54	Deflazacort	83.75
Trimetrexate	53.36	Alizapride	66.38	Ouabain	83.12
Fenoldopam	53.30	Triptorelin	66.24	Proscillaridin	82.75
Carvedilol	52.94	Gonadorelin	66.19	Digitoxin	82.15
Emetine	52.78	Histrelin	65.93	Hydrocortisone	81.96
Methotrimeprazine	52.66	Leuprolide	65.81	Deslanoside	81.87

Amodiaquine MS		Astemizole MS		Budesonide MS	
Amodiaquine	100.00	Astemizole	100.00	Budesonide	100.00
Quinacrine	67.53	Pimozide	77.30	Flurandrenolide	91.67
Chloroquine	63.24	Droperidol	72.98	Flunisolide	91.31
Hydroxychloroquine	61.24	Domperidone	72.94	Triamcinolone	90.29
Glafenine	59.25	Puromycin	67.69	Fluocinolone	89.43
Primaquine	59.14	Clozapine	66.54	Halcinonide	87.19
Hycanthone	56.67	Loxapine	64.81	Fluocinonide	86.06
Methylcoclaurine	55.44	Kinetin	64.03	Cymarin	85.13
Ethaverine	54.42	Pirenzepine	63.93	Amcinonide	84.50
Clebopride	54.14	Prazosin	63.80	Hydrocortisone	83.95
Nomifensin	53.57	Amoxepine	63.74	Peruvoside	83.76
Laudanosine	53.53	Deoxyadenosin	63.64	Convallatoxin	83.52
Papaverine	53.39	Pipemidic acid	62.15	Strophantin	83.27
Berbamine	53.08	Trazodone	62.03	Ouabain	83.12
Tetrahydrosappanone	53.00	Omeprazole	62.00	Periplocymarin	83.12
Deshydrosappanol	53.00	Enoxacin	61.67	Hydroxycarapin	82.99
Emetine	52.78	Vidarabine	61.63	Gitoxin	82.73
Homidium	52.41	Adenosine	61.63	Oleandrin	82.40
Dequalinium	52.34	Lansoprazole	61.45	Deoxyandirobin	82.33
Hydroxytacrine	52.08	Urapidil	61.18	Gedunin	82.20

Forts. Tab. 2.1. *Blau: inaktiv in Kociskos Assay; Fett: IC_{50} von $\leq 10 \mu M$ in Kociskos Assay; Fett grün: IC_{50} von $\leq 1 \mu M$ in Kociskos Assay; Fett rot: als neue Leitstruktur vorgeschlagen*

Clomiphene SD		Lovastatin SD		Prochlorperazine SD	
Clomiphene	100.00	Lovastatin	100.00	Prochlorperazine	100.00
Tamoxifen	74.68	Simvastatin	94.06	Ethylthiopazine	93.90
Toremifene	73.75	Pravastatin	82.46	Perazine	92.54
Etafenone	58.11	Ciclesonide	63.43	Perphenazine	91.91
Chlorotrianisene	56.80	Gitoformate	63.42	Butaperazine	88.32
Biphenamine	48.32	Pengitoxin	63.37	Thiopropazate	87.44
Chlorphenoxamine	46.92	Peruvoside	62.83	Chlorpromazine	85.75
Amiodarone	46.92	Acetyldigitoxin	62.82	Thioridazine	85.21
Procaine	45.80	Fungichromin	62.64	Chlorproethazine	84.82
Diphenhydramine	45.71	Proscillaridin	62.59	Acetophenazine	83.17
Ebastine	45.65	Cymarine	62.52	Periciazine	81.02
Phenoxybenzamine	45.59	Isolanid	62.12	Trifluoperazine	80.69
Tolterodine	44.55	Mupirocin	62.11	Dixyrazine	80.00
Benzethonium	43.55	Hydrocortisone	62.04	Methdilazine	79.95
Gallamine	43.19	Ivermectin	61.97	Mesoridazine	79.25
Carbinoxamine	43.12	Candicidin	61.80	Promazine	76.08
Orphenadrine	42.75	Digitoxin	61.76	Acepromazine	75.93
Bibenzonium	42.64	Acetyldigoxin	61.72	Mequitazine	75.91
Eprazinone	42.61	Methylprednisolone	61.63	Fluphenazine	75.00
Cloperastine	42.60	Carbenoxolone	61.63	Prothipendyl	73.30

Clomiphene MS		Lovastatin MS		Prochlorperazine MS	
Clomiphene	100.00	Lovastatin	100.00	Prochlorperazine	100.00
Tamoxifen	74.68	Compactin	97.53	Perphenazine	91.91
Chlorotrianisene	56.80	Deoxyandiobin	67.74	Chlorpromazine	85.75
Adiphenine	49.81	Nymania	67.33	Thioridazine	85.21
Tilorone	47.34	Khayasin	67.05	Periciazine	81.02
Amiodarone	46.92	Crassin	66.89	Trifluoperazine	80.69
Procaine	45.80	Mexicanolide*	66.86	Promazine	76.08
Phenoxybenzamine	45.59	Mexicanolide*	66.09	Flufenazine	75.00
Proadifen	44.82	Mexicanolide*	66.02	Ethiopropazine	73.18
Benzethonium	43.13	Mexicanolide*	66.02	Methiothepin	73.00
Carbinoxamine	43.12	Mexicanolide*	66.01	Promethazine	71.61
Orphenadrine	42.75	Mexicanolide*	65.89	Propiomazine	70.79
Methylbenzthonium	42.65	Mexicanolide*	65.88	Trimeprazine	70.66
Cloperastine	42.60	Hydroxycarapin	65.71	Triflupromazine	69.51
Pyrilamine	42.43	Khayasin C	65.70	Isothipendyl	66.74
Drofenine	42.42	Oxokhivorin	65.54	Clozapine	66.31
Bepridril	42.06	Carapin-8(9)-ene	65.43	Tetrandrine	65.38
GBR 12909	42.02	Carapin	65.43	Clomipramine	64.64
Hydroxyzine	42.01	Gedunin	65.37	Mianserin	64.32
Mebeverine	41.53	Loganin	64.77	Berbamine	64.14

Forts. Tab. 2.1. *Blau: inaktiv in Kociskos Assay; Fett: IC_{50} von $\leq 10 \mu M$ in Kociskos Assay; Fett grün: IC_{50} von $\leq 1 \mu M$ in Kociskos Assay; Fett rot: als neue Leitstruktur vorgeschlagen*

Quinacrine SD		Terfenadine SD		Thioridazine SD	
Quinacrine	100.00	Terfenadine	100.00	Thioridazine	100.00
Chloroquine	72.46	Fexofenadine	82.12	Mesoridazine	92.63
Hydroxychloroquine	69.61	Ebastine	75.87	Ethylthioperazine	88.86
Ethacridine	68.18	Bucizine	67.34	Prochlorperazine	85.21
Amodiaquine	67.53	Moperone	61.89	Perazine	82.56
Dequalinium	64.45	Bromperidol	59.94	Butaperazine	81.10
Chlorproethazine	64.44	Haloperidol	59.94	Chlorproethazine	80.25
Primaquine	64.01	Trihexyphenidyl	59.57	Perphenazine	79.86
Emetine	62.72	Loperamide	59.02	Chlorpromazine	79.64
Methotrimeprazine	61.52	Pridinol	58.70	Methdilazine	78.75
Hydroquinine	61.40	Budipine	58.13	Thiopropazate	77.98
Carvedilol	60.90	Procyclidine	57.65	Periciazine	76.92
Chlorpromazine	60.48	Ifenprodil	57.24	Acetophenazine	76.91
Meprotixol	60.45	Hydroxyzine	56.59	Mequitazine	76.56
Dehydroemetine	59.39	Meclizine	56.23	Acepromazine	75.73
Gefitinib	59.08	Difenoxin	55.97	Ethiopropazine	75.19
Thiopropazate	58.97	Phenoperidine	55.20	Dixyrazine	75.12
Prochlorperazine	58.88	Clemastine	55.12	Trifluoperazine	74.51
Perphenazine	58.50	Flunarizine	54.43	Promazine	74.03
Alfuzosin	58.29	Cinnarizine	54.39	Methotrimeprazine	72.88

Quinacrine MS		Terfenadine MS		Thioridazine MS	
Quinacrine	100.00	Terfenadine	100.00	Thioridazine	100.00
Chloroquine	72.46	Fexofenadine	82.12	Prochlorperazine	85.21
Hydroxychloroquine	69.61	Haloperidol	59.94	Perphenazine	79.86
Amodiaquine	64.87	Trihexyphenidyl	59.57	Chlorpromazine	79.64
Primaquine	64.01	Loperamide	59.02	Periciazine	76.92
Emetine	62.72	Pridinol	58.70	Methiothepin	75.87
Dequalinium	62.06	GBR 12909	58.28	Ethiopropazine	75.19
Hydroxyquinidine	61.40	Procyclidine	57.65	Trifluoperazine	74.51
Hydroxytetracine	61.24	Hydroxyzine	56.59	Promazine	74.03
Berbamine	61.23	Meclizine	56.23	Promethazine	72.35
Hycanthone	60.91	Clemastine	55.12	Propiomazine	72.29
Chlorpromazine	60.48	Cinnarizine	54.88	Trimeprazine	70.15
Isocorydine	59.73	Flunarizine	54.43	Flufenazine	70.08
Tetrahydropalmatine	59.18	Cloperastine	53.21	Trifluopromazine	69.45
Boldine	59.01	Lobelanidine	53.14	Isothipendyl	67.48
Tetrandrine	58.99	Cyclizine	51.57	Tetrandrine	66.45
Prochlorperazine	58.88	Lobeline	49.24	Berbamine	65.90
Perphenazine	58.50	Diphenylpyraline	49.12	Emetine	61.93
Quinine	57.71	Vesamicol	48.81	Clozapine	61.91
Quinidine	57.71	Venlafaxine	48.62	Mianserin	61.20

Forts. Tab. 2.1. *Blau: inaktiv in Kociskos Assay; Fett: IC_{50} von $\leq 10 \mu M$ in Kociskos Assay; Fett grün: IC_{50} von $\leq 1 \mu M$ in Kociskos Assay; Fett rot: als neue Leitstruktur vorgeschlagen*

Thiothixene SD		Trifluoperazine SD	
Thiothixene	100.00	Trifluoperazine	100.00
Thiopropazine	84.73	Fluphenazine	93.26
Pipothiazine	80.82	Triflupromazine	88.05
Fonazine	78.62	Prochlorperazine	80.69
Metopimazine	73.10	Butaperazine	80.27
Flupenthixol	69.37	Perazine	79.01
Oxomemazine	66.61	Ethylthiopropazine	79.01
Sildenafil	65.89	Acetophenazine	75.97
Delavirdine	65.18	Perphenazine	75.16
Argatroban	65.02	Flupenthixol	75.00
Vardenafil	64.91	Thiopropazine	75.00
Gliquidone	64.87	Thiopropazine	74.73
Trifluoperazine	64.77	Thioridazine	74.51
Polythiazide	63.91	Mesoridazine	72.05
Bietaserpine	62.84	Periciazine	71.86
Methyclothiazide	62.75	Pipothiazine	71.22
Quinagolide	62.60	Dixyrazine	70.94
Zafirlukast	62.57	Metopimazine	70.85
Butaperazine	62.13	Methdilazine	70.47
Mesoridazine	62.03	Chlorprothazine	70.29

Thiothixene MS		Trifluoperazine MS	
Thiothixene	100.00	Trifluoperazine	100.00
Sildenafil	65.89	Flufenazine	93.26
Trifluoperazine	64.77	Triflupromazine	88.05
Flufenazine	61.72	Prochlorperazine	80.69
Cyclothiazide	61.68	Perphenazine	75.16
Nicergoline	61.31	Thioridazine	74.51
Propiomazine	61.00	Periciazine	71.86
Althiazide	60.81	Chlorpromazine	69.59
Vinpocetine	60.38	Propiomazine	66.25
Metolazone	60.30	Ethiopropazine	65.86
Gambogic acid	60.06	Promazine	65.59
Benzthiazide	60.03	Thiothixene	64.77
Ofloxacin	59.91	Methiothepin	63.91
Bendroflumethiazide	59.89	Promethazine	63.74
Aclacinomycin	59.85	Isotipendyl	63.39
Triflupromazine	59.71	Trimeprazine	62.44
Rifampin	59.41	Clozapine	61.92
Aklavine	59.40	Mefloquine	60.90
Rutilantin	59.40	Tetrandrine	60.74
Meloxicam	59.10	Loxapine	60.00

2.7 3D-Struktursuche

Wie oben erläutert, führt die 2D-Suche in Strukturdatenbanken meist zu Komponenten mit großer chemischer Ähnlichkeit zu den Ausgangssubstanzen. Strukturen aus anderen chemischen Klassen (sog. *scaffold hopper*) können hingegen eher durch strukturelle Überlagerungen (s. S. 8) aufgefunden werden. Der Algorithmus beachtet die Anordnung der einzelnen Atome im Raum. Ein großer Vorteil bei dieser Herangehensweise ist, daß strukturelle Ähnlichkeiten von chemisch „unähnlichen“ Substanzen aufgespürt werden, die durch 2D-Vergleiche nicht entdeckt werden können.

Exemplarisch ist in Abb. 2.3 eine Überlagerung von Prochlorperazin mit Opipramol gezeigt. Obwohl die Substanzen zu verschiedenen chemischen Klassen gehören und deren Ähnlichkeit nach Tanimoto gering ist (63%), findet der Überlagerungsalgorithmus strukturelle Verwandtschaft. Dabei wird ein Sechsring im Prochlorperazin mit einem Siebenring im Opipramol überlagert, die Elektronendichte eines Schwefelatoms ist durch eine Doppelbindung ersetzt. Wie Prochlorperazin enthält auch Opipramol einen Piperazinring. Auch die Ringsysteme von Clopenthixol, Methanthelin und Rupatadin (Abb. 2.4) wurden durch strukturelle Überlagerung mit Prochlorperazin gefunden.

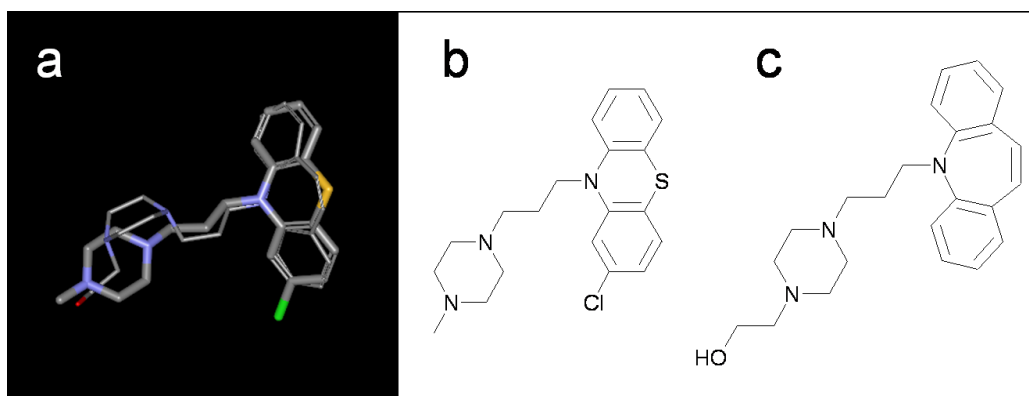


Abbildung 2.3: Strukturelle Überlagerung zwischen Prochlorperazin und Opipramol (a). Ein Sechsring im Prochlorperazin (b) wird mit einem Siebenring des Opipramol (c) überlagert.

2.8 16 neue potentielle Inhibitoren

Suchergebnisse der 2D- und 3D-Suche wurden zusammengestellt und nach strukturellen Gesichtspunkten visuell mit aktiven und inaktiven Substanzen verglichen. Durch Ausschluß von Substanzen mit großen strukturellen Ähnlichkeiten zu inaktiven Substanzen wurde die Liste der interessanten Wirkstoffe systematisch verkleinert. Ebenso wurde im Falle von hoher Ähnlichkeit zwischen mehreren Substanzen nur jeweils ein Vertreter der Klasse in die Liste übernommen.

Abb. 2.4 zeigt die verbleibende Liste von 16 Substanzen, die als potentielle neue Leitstrukturen in der TSE-Therapie betrachtet werden. Ein großer Teil der Substanzen wurden durch Ähnlichkeitssuchen mit Phenothiazinen gefunden. Das Antihistaminikum Mizolastin ist dem Astemizol aus der gleichen Wirkstoffklasse ähnlich. Eine Verschiebung der Indikation ist zum Beispiel beim antineoplastisch wirkenden Toremifen zu beobachten, das aufgrund seiner Ähnlichkeit zum Ovulationsstimulans Clomiphen gefunden wurde. Buclicin und Ebastin zeigen Ähnlichkeiten zu dem Antihistaminikum Terfenadin, das Antihistaminikum Rupatadin wurde durch 3D-Suchen mit Quinacrin aufgefunden.

Zur Zeit werden die Substanzen aus der Liste von experimentellen Gruppen geprüft. Das Spektrum an Substanzen reicht von Komponenten mit hoher struktureller Ähnlichkeit zu bereits bekannten Inhibitoren wie z. B. Perazin bis zu komplett neuen Leitstrukturen wie Opipramol oder Oxomemazin. Die therapeutische Klassifizierung der aufgefundenen Wirkstoffe weicht oft von der Klassifizierung der von Kocisko untersuchten Substanzen ab – die Methode erlaubt also auch das Auffinden von Strukturen aus neuen chemischen Stoffklassen (*scaffold hoppers*).

2.9 Diskussion

Ausgehend von publizierten Inhibitoren der Fehlfaltung von PrP konnten durch 2D- und 3D-Suche in der SuperDrug-Datenbank weitere Substanzen als potentielle Inhibitoren vorgeschlagen werden. Als Kriterium dient dabei nicht nur die Ähnlichkeit zu bekannten Inhibitoren im positiven Sinne, sondern auch große Ähnlichkeit zu inaktiven Substanzen als Ausschlußkriterium.

Es konnte gezeigt werden, daß Substanzen mit hoher Ähnlichkeit zu Inhibitoren oft auch selbst inhibitorisch wirken. Dabei gibt es offensichtlich einen Unterschied zwischen Substanzen verschiedener chemischer Klassen: Während Phenothiazine zum Großteil zumindest eine geringe inhibitorische Wirkung besitzen, wurde unter den Steroiden nur Budenosid identifiziert –

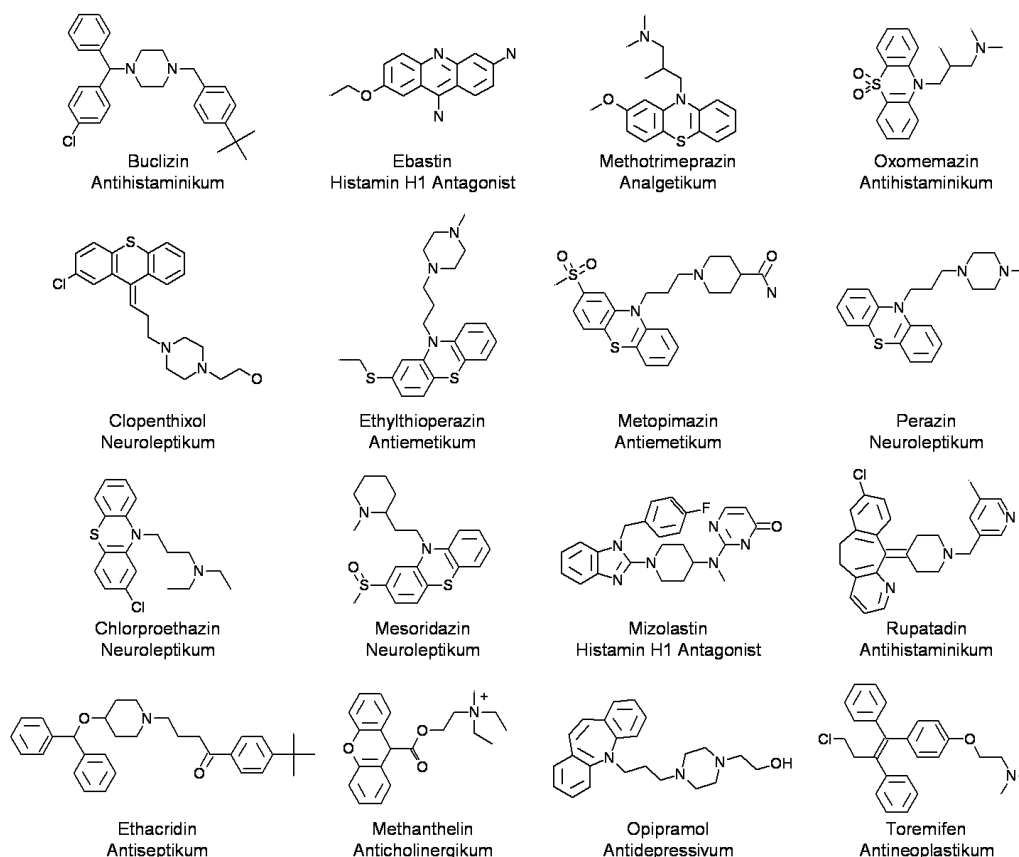


Abbildung 2.4: Neue potentielle Inhibitoren der PrP^{Sc} -Bildung.

anderen Steroide besitzen offensichtlich keine Wirkung.

Für die Wirkung der Phenothiazine ist demnach offensichtlich ihr Ringsystem verantwortlich. Durch Vergleiche von aktiven und inaktiven Substanzen konnten weitere Bedingungen für eine inhibitorische Wirkung gefunden werden. So sind zum Beispiel polare Reste am Piperazinring scheinbar für eine Interaktion ungünstig.

Während die 2D-Struktursuche unter Verwendung des Tanimoto-Koeffizienten hauptsächlich Substanzen aus der gleichen chemischen Klasse wie das Ursprungsmolekül findet, bietet eine 3D-Suche die Möglichkeit, auch Verbindungen anderer Substanzklassen, sog. *scaffold hopper*, zu finden. In der Liste der vorgeschlagenen Inhibitoren finden sich einige Beispiele für derartige Substanzen (z. B. Opipramol, Oxomemazin).

Die letztendlich vorgeschlagenen Substanzen zeigen eine große strukturelle und pharmakologische Vielfalt. Besonders durch 3D-Suchen konnten auch

Substanzen aufgefunden werden, deren Ähnlichkeit zu den Ausgangsstrukturen durch 2D-Suchen nicht festgestellt werden könnte.

Durch die *in silico* Suche nach Ähnlichkeit zu bekannten Inhibitoren konnte aus den 2 300 in der SuperDrug abgelegten Strukturen ein überschaubares Set von 16 Substanzen generiert werden, deren Effektivität in der Inhibition der PrP^{Sc}-Bildung zur Zeit experimentell überprüft wird. Durch die Verkleinerung des Suchraumes ist es somit möglich, die aufgefundenen Substanzen intensiv zu testen, was zweifellos bei 2 300 Substanzen nicht möglich wäre.

Ein weiterer Vorteil besteht darin, daß die aufgefundenen Substanzen gängige pharmazeutische Wirkstoffe sind. Daher entfallen bei der eventuellen späteren Verwendung im therapeutischen Bereich zeit- und kostenintensive klinische Studien zur Toxizität und Pharmakokinetik. Da die Substanzen in der Datenbank gut charakterisierten pharmakologischen Wirkstoffklassen entstammen, stehen bei Erfolg zu jeder Substanz noch weitere, ähnliche Substanzen zur Verfügung.

Kapitel 3

Ähnlichkeiten zwischen Sequenzmotiven und ihre funktionelle Bedeutung

3.1 Peroxisomaler Proteinimport

3.1.1 Peroxisomen haben wichtige Aufgaben im Stoffwechsel

Im Elektronenmikroskop erscheinen Peroxisomen als elektronendichte, von einer einfachen Membran umgebene Strukturen, die oft von großen Proteinkristallen ausgefüllt sind. Ihr Durchmesser beträgt 0,1–1 μm . Der Name „Peroxisom“ [96] deutet auf Peroxide hin, die in diesem Organell auftreten. In der Tat waren die ersten bekannten peroxisomalen Enzyme Oxidasen, die Wasserstoffperoxid erzeugen, und Katalase, die es in Wasser und Sauerstoff spaltet [97].

Lange Zeit hielt man den Abbau giftiger Peroxide für die einzige Aufgabe der Peroxisomen. Heute wissen wir, daß Peroxisomen wichtige Funktionen in diversen Stoffwechselwegen erfüllen (zur Übersicht siehe [98]): Während im Menschen nur lang- und verzweigtkettige Fettsäuren peroxisomal abgebaut werden [99], bauen niedere Pilze Fettsäuren ausschließlich in Peroxisomen ab [100].

Auch in der Cholesterin- und Gallensäurebiosynthese [101, 102], der Etherphospholipidsynthese [103], dem Prostaglandinstoffwechsel [104] und dem Polyamin- und Purinmetabolismus [105, 106] spielen Peroxisomen eine wichtige Rolle. Weiterhin sind Peroxisomen an der Penicillinbiosynthese in niederen Pilzen beteiligt [107].

Mutationen in peroxisomalen Proteinen sind im Menschen mit Krankheiten, den *Peroxisomal Biogenesis Disorders* (PBDs), assoziiert (zur Übersicht siehe [108, 109, 110]). Erste Hinweise auf einen Zusammenhang zwischen dem Zellweger-Syndrom und Peroxisomen lagen bereits 1973 vor [111]. Die Zellweger-Krankheit äußert sich, wie auch verwandte PBDs, in schwerwiegenden neurologischen Störungen sowie Leber- und Niereninsuffizienz, was meist zum Tod innerhalb der ersten Lebensjahre führt. Eine andere, durch den Film *Lorenzos Öl* ins öffentliche Bewußtsein gerückte PBD ist die Adrenoleukodystrophie (ALD).

Peroxisomale Krankheiten unterteilt man in drei Hauptgruppen [112]: Bei Störungen der ersten Gruppe liegt ein genereller peroxisomaler Funktionsverlust vor, man findet jedoch üblicherweise noch leere peroxisomale Membranstrukturen, die sogenannten *ghosts*, die auch peroxisomale Membranproteine (PMPs) enthalten. Zu dieser Gruppe gehören das Zellweger-Syndrom [113] und die neonatale Adrenoleukodystrophie.

Krankheiten der zweiten Gruppe zeigen einen partiellen Importdefekt, der nur einige peroxisomale Proteine betrifft, wie zum Beispiel die Rhizomelische Chondrodysplasia punctata [114, 115, 116]. Bei Krankheiten der dritten Gruppe ist nur ein einzelnes peroxisomales Enzym defekt, der Import an sich funktioniert aber. Zu dieser Gruppe zählen zum Beispiel die X-chromosomale Adrenoleukodystrophie, das Pseudozellweger-Syndrom und die Refsum-Krankheit [117].

3.1.2 Import peroxisomaler Matrixproteine

Peroxisomale Proteine werden an freien Ribosomen im Cytosol synthetisiert und posttranslational importiert [118]. Matrixproteine, die importiert werden sollen, enthalten eines der beiden peroxisomalen Targetingsignale (PTS): Die C-terminale Erkennungssequenz [SAC][HKR][LM] (PTS1 [119]) oder die Sequenz [RK][LIV]X₅[HQ][LA] (PTS2 [120]) am N-Terminus.

Die Konsensussequenzen der Targetingsignale sind in unterschiedlichsten Organismen sehr ähnlich und stark konserviert [119]. Auch Proteine, die nicht peroxisomalen Ursprungs sind, können durch Addition eines PTS importiert werden, diese Sequenzen sind daher notwendig und hinreichend für den peroxisomalen Import [121].

Proteine beider Untergruppen binden nach der Synthese im Cytosol einen löslichen Rezeptor: Pex5 [122] für PTS1-Proteine und Pex7 [123, 124] für PTS2-Proteine. Der Rezeptor-Cargo-Komplex gelangt nun zum Peroxisom, wo er an den Dockingkomplex bindet und dann in das Organell importiert wird.

Die PTS-Konsensussequenzen konnten in weiteren Studien [125, 126] genauer beschrieben werden, für PTS1-Sequenzen existiert ein Vorhersage-Algorithmus [127]

3.1.3 Import peroxisomaler Membranproteine

Am Import peroxisomaler Membranproteine sind andere Komponenten beteiligt. Hettema *et al.* [128] identifizierten als die einzigen Peroxine, deren Verlust in *S. cerevisiae* auch peroxisomale Reststrukturen, die sogenannten *ghosts*, verschwinden läßt, Pex3 und Pex19. In anderen Organismen ist zur Ausbildung peroxisomaler Membranen auch Pex16 nötig [129], das in *S. cerevisiae* nicht bekannt ist.

Da Pex19 das einzige der beteiligten Proteine ist, das cytosolisch vorkommt [130] (Pex3 und Pex16 sind peroxisomale Membranproteine), wird davon ausgegangen, daß Pex19 der cytosolische Rezeptor für peroxisomale Membranproteine darstellt, die posttranslational importiert werden.

Durch Deletionsstudien wurden bereits für diverse peroxisomale Membranproteine die fürs Targeting relevanten Bereiche ermittelt (Tab. 3.1). Es fällt auf, daß diese Bereiche meist größere Abschnitte der beteiligten Proteine umfassen, teilweise (Pex13, Pex16) sind die Informationen auch widersprüchlich oder sprechen für mehrere für das Targeting verantwortliche Bereiche im Protein.

Gemeinsam ist den Targetingsignalen die Notwendigkeit von mindestens einer Transmembrandomäne in Zusammenhang mit einer kurzen Sequenz, die basische und hydrophobe Aminosäuren enthält. Eine Konsensussequenz konnte hingegen bisher nicht gefunden werden.

Auffällige Ähnlichkeit mit den publizierten Targeting-Sequenzen finden sich zu Bereichen, von denen eine Pex19-Interaktion gezeigt wurde (Tab. 3.2). Obwohl nicht in allen Fällen konsistent, kann eine Rolle von Pex19 bei der Insertion peroxisomaler Membranproteine daher angenommen werden.

Es wurde auch vorgeschlagen, daß Pex19 als Chaperon die Transmembrandomänen peroxisomaler Membranproteine unspezifisch bindet, um eine Aggregation oder Fehlfaltung zu verhindern [135].

Um die Rolle von Pex19 beim peroxisomalen Membranprotein-Import zu verstehen ist es daher erforderlich, die Pex19-Interaktion peroxisomaler Membranproteine experimentell und bioinformatisch genauer zu untersuchen.

Tabelle 3.1: publizierte Targeting-Bereiche einiger peroxisomaler Membranproteine.

PMP	targeting region
Pex2	130–159+TMD1 [131]
Pex11 β	181–259 [132]
Pex12	50–233 [133]
Pex13	145–233; 1–174; 179–403 [133]; 1–193; 194–403 [134]
PEX14	1–147 [132]
PEX16	59–219 [133]; 221–336 [135]; 70–140 [136]
PMP22	1–94; 95–195 [137]
PMP34	1–147; 244–307; [134]
ALDP	1–281 [138]; 67–164 [139]
PMP69	16–113 [139]
PMP70	61–160 [140]; 1–124 [132]

Tabelle 3.2: publizierte Pex19-Bindestellen einiger peroxisomaler Membranproteine

PMP	binding region
PEX11 β	181–259; 211–259 [132]
PEX12	275–359 [133]
PEX13	145–233 [133]
PEX14	22–81 [141]; 1–147 [132]
PEX16	59–219 [133]; 59–219; 221–336 [135]
PMP22	1–94; 95–195 [135]; 75–174 [137]
PMP34	1–147; 244–307 [134]; 270–307 [135]
ALDP	66–203 [138]
ALDRP	1–218 [138]
PMP70	60–160; 80–160 [140]; 1–61; 1–81 [132]; 1–61; 61–160 [135]

3.2 Zielsetzung

In Zusammenarbeit mit experimentellen Arbeitsgruppen soll die Interaktion von Pex19 mit peroxisomalen Membranproteinen charakterisiert und Bereiche von Proteinen, die mit Pex19 interagieren untersucht werden. Die Pex19-Bindestellen weiterer Membranproteine sollen vorhergesagt werden. Die im Folgenden dargestellten experimentellen Ergebnisse wurden nicht vom Autor erhoben, sondern entstammen der Zusammenarbeit mit experimentellen Arbeitsgruppen. Peptidspot-Experimente wurden im Labor von Rudolf Volkmer-Engert durchgeführt, Yeast-Two-Hybrid Experimente im Labor von Ralf Erdmann.

3.3 Peptide aus Pex13 interagieren mit Pex19

Einer der wichtigsten Bindungspartner von Pex19 ist Pex13. Um den Bereich des Proteins ausfindig zu machen, der mit Pex19 interagiert, wurden im Labor von Rudolf Volkmer-Engert Peptidfragmente aus Pex13 auf CAPE-Cellulosemembranen [142] mit Hilfe der SPOT-Synthese [143] synthetisiert (Abb. 3.1). Dabei werden nach dem Merrifield-Verfahren Peptide einheitlicher Länge auf Cellulosemembranen synthetisiert. Die Peptide sind mittels eines Ankers aus zwei β -Alaninresten an die Matrix gekoppelt, der N-Terminus bleibt frei.

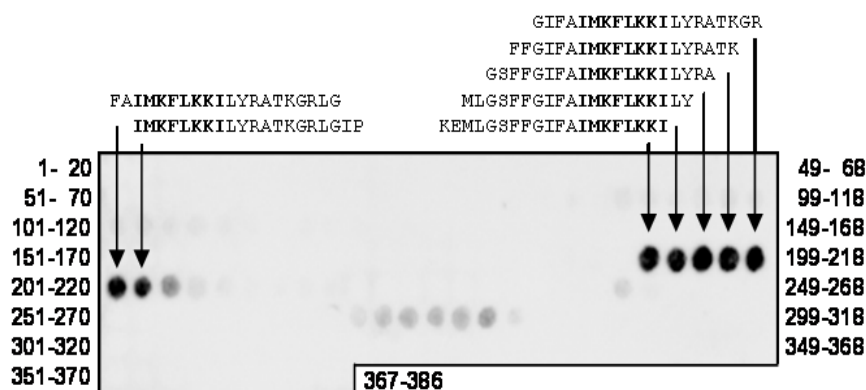


Abbildung 3.1: Peptidscan von Pex13. Es ist deutlich eine Binderegion zu erkennen.

Hier wurden abschnittsweise je 20 Aminosäuren lange Peptide mit einem Versatz von zwei Aminosäuren synthetisiert, die das gesamte Protein abdecken. Man erkennt deutlich einen Bindebereich mit der zentralen Sequenz

IMKFLKKI. Die Region liegt zwischen den beiden Transmembrandomänen von Pex13.

Im Labor von Ralf Erdmann konnte auch mittels *Yeast Two Hybrid* Technik durch Verkürzungskonstrukte gezeigt werden, daß der Bereich von Aminosäure 173 bis 233 aus Pex13 für eine Interaktion mit Pex19 hinreichend ist (Abb. 3.2).

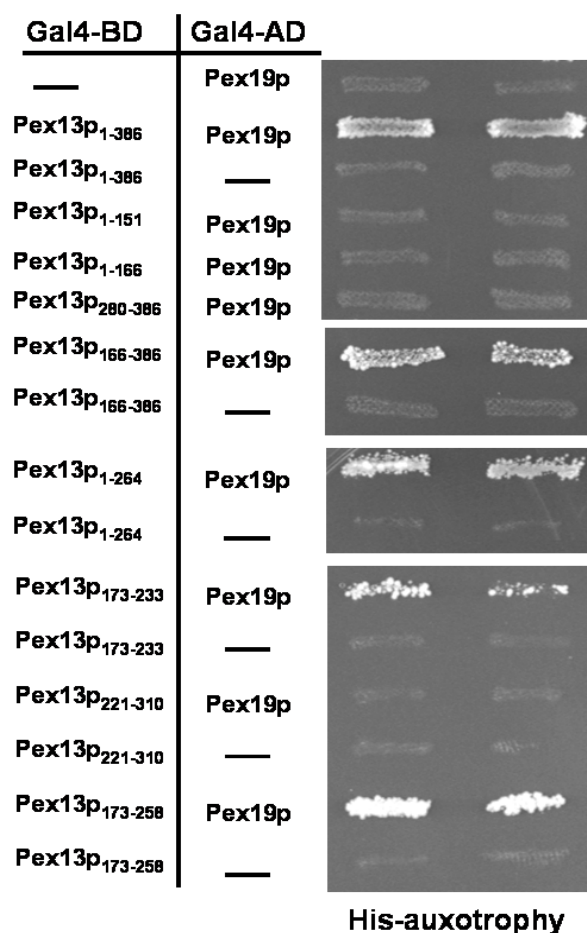


Abbildung 3.2: *Yeast-Two Hybrid* Analyse der Bindung von Pex13 an Pex19. Der Bereich von Aminosäure 173 bis 233 ist für die Bindung ausreichend.

Der aufgefundene Bereich der Pex19-Interaktion liegt innerhalb der beschriebenen zur peroxisomalen Lokalisation und Pex19-Bindung notwendigen Bereiches von Aminosäure 145 bis 233, und Punktmutationen, die die Pex19-Interaktion verhindern, konzentrieren sich auf den hier gefundenen Bereich [133].

Wie lang muß ein Peptidfragment sein, um mit Pex19 zu interagieren? Die auf der Membran in Abb. 3.1 synthetisierten Peptide waren je 20 Aminosäuren lang. Als nächstes wurden mittels SPOT-Synthese systematisch Verkürzungskonstrukte im Bereich der interagierenden Aminosäuren synthetisiert (Abb. 3.3).

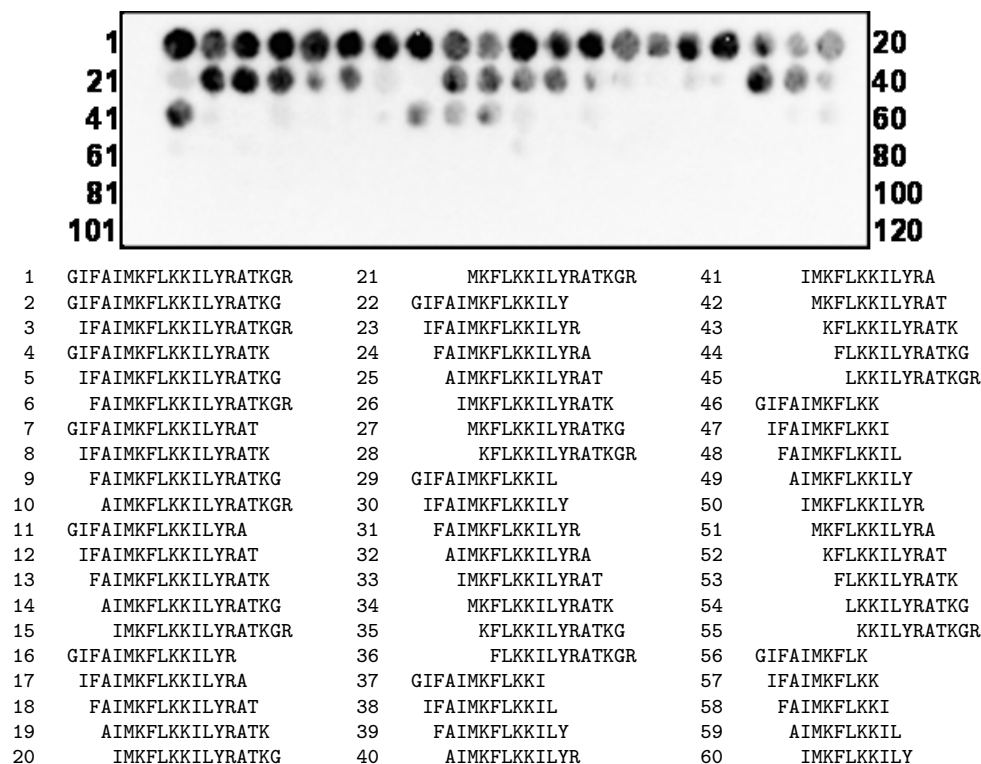


Abbildung 3.3: Längenanalyse des bindenden Bereiches aus Pex13p. Auf der Membran sind Verkürzungskonstrukte der in Abb. 3.1 aufgefundenen Binderegion aufgetragen.

Die kürzesten Peptide mit detektierbarer Pex19-Bindung sind die auf den Spots 48–50 zu sehenden 11-mere, die die Sequenz IMKFLKKIL enthalten. Das zentrale Stück mit neun Aminosäuren Länge entspricht dem gemeinsamen Sequenzabschnitt der interagierenden Peptide aus Abb. 3.1.

Die neun Aminosäuren lange Core-Sequenz selbst bindet allerdings nicht an Pex19. Offensichtlich ist für die Interaktion eine Proteinumgebung der Sequenz nötig, die fehlende Interaktion könnte auch ein Artefakt aufgrund der Spotsynthese auf der Membran darstellen. Weiterhin ist möglich, daß der freie Terminus des auf der Membran befindlichen Peptids die Interaktion stört.

Die Zusammensetzung des Peptides entspricht den publizierten Motiven

aus basischen (K) und hydrophoben (I,F,L) Aminosäuren. Um näher zu untersuchen, welche Aminosäuren für die Interaktion essentiell sind, wurde eine Mutationsanalyse durchgeführt (Abb. 3.4).

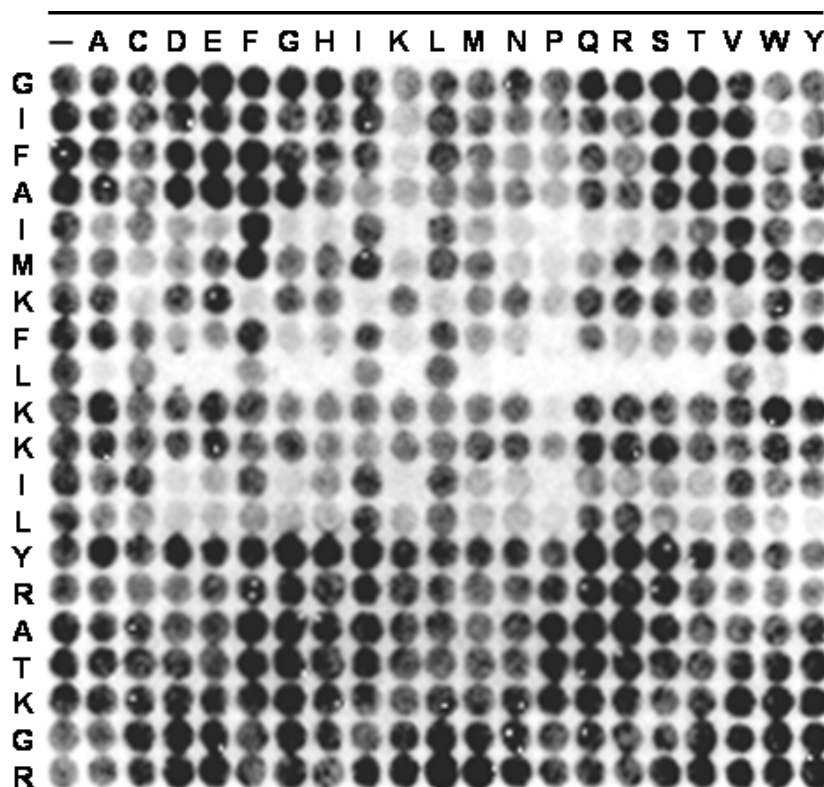


Abbildung 3.4: Mutationsanalyse des 20 Aminosäuren langen Bindebereichs aus Pex13p. Jede der 20 Aminosäuren wurde sukzessive durch alle anderen Aminosäuren ausgetauscht.

Dabei werden sukzessive die Aminosäuren an allen Positionen des zu untersuchenden Peptids durch alle 20 Aminosäuren ausgetauscht und die resultierenden Peptide auf ihre Pex19-Bindung überprüft.

Während Aminosäureaustausche an den beiden Enden des Peptides keine weitreichenden Auswirkungen zeigen und die Bindung an Pex19 erhalten bleibt, ist der schon in den vorherigen Analysen aufgefundene mittlere Bereich kritisch. Insbesondere der Leucinrest an Position 9 kann nur durch ähnliche Aminosäuren ausgetauscht werden.

An der Abbildung wird ein deutlicher Mangel der Technik sichtbar: die Meßwerte streuen erheblich. So ist z. B. auf der in Abb. 3.4 das Originalpeptid aus Pex13 40mal enthalten. Die Meßwerte streuen zwischen ca. 15 000 und 50 000. Dies ist sowohl auf unterschiedliche Effizienz des Synthese als

auch auf ungleichmäßige Benetzung der Oberfläche, $\frac{1}{2}$ Å durch Antikörperlösung beim Entwickeln und schließlich Ungenauigkeiten beim Auslesen der Phosphoreszenzintensitäten zurückzuführen [144]. Ein weiteres Problem besteht darin, daß die Spots nicht immer kreisrund sind und daher teilweise die Hintergrundintensität als Teil des Spots gezählt wurde oder Teile des Spots außerhalb der Erfassungsgrenzen des Spots durch die Software liegen.

Durch *Yeast Two Hybrid* Technik konnte im Labor von Ralf Erdmann gezeigt werden, daß auch im Kontext des *full length* Proteins eine Mutation des Leucinrestes an Position 9 die Pex19-Interaktion verhindert [145].

3.4 Pattern-Suchen

Aus der in Abb. 3.4 gezeigten Mutationsanalyse können die an den jeweiligen Aminosäurepositionen tolerierten Aminosäuren abgeleitet werden. Dies führt zu einem *Pattern* unter Tolerierung der Aminosäuren, die in der Mutationsanalyse noch eine Interaktion mit Pex19 ermöglichten. Als restriktivste Position erscheint hier die Aminosäure an Position 9, die randständigen Aminosäuren sind offensichtlich für die Bindung nicht kritisch.

Aus den Ergebnissen der Mutationsanalyse wurde das Pattern XXXX[CFILTVW]XX[ACFILQVWY][CILV]XX[ACFILVWY][ILQRV]XX herausgearbeitet. Um möglichst alle Bindestellen bei einer Suche im Proteom aufzufinden, wurde das Pattern möglichst nichtrestriktiv gestaltet. Dies beinhaltet allerdings den Nachteil, daß eine proteomweite Suche zu sehr vielen Treffern führt.

Um die Suche einzugrenzen, wurden nur peroxisomale Membranproteine und als Kontrollgruppe Proteine der Vakuolenmembran aus *S. cerevisiae* durchsucht. Die Suche führte zu 119 Treffern, die ebenfalls auf einer Cellulosemembran synthetisiert und auf Pex19-Interaktion getestet wurden. Abb. 3.5 zeigt die Interaktion der Peptide aus peroxisomalen Membranproteinen.

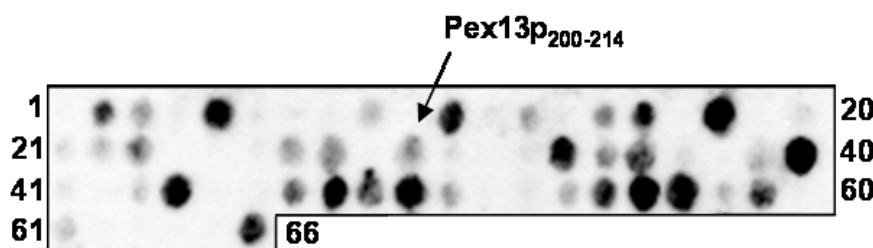


Abbildung 3.5: Peptid-Blot der Pattern-Hits zur Überprüfung der Interaktion mit Pex19

Obwohl das gewählte Pattern unrestrictiv ist, zeigen viele der getesteten Peptide eine Interaktion mit Pex19. Unter den aufgefundenen Peptiden befindet sich auch das Ursprungspeptid aus Pex13. Auffällig ist, daß viele Peptide eine noch deutlich stärkere Interaktion zeigen.

Entgegen den Erwartungen interagieren auch teilweise Peptide, in denen die Aminosäure an Stelle 9 durch Alanin ausgetauscht ist, so daß die aus der Mutationsanalyse des Pex13-Peptids abgeleitete Notwendigkeit einer hydrophoben Aminosäure an dieser Position nicht bestätigt werden konnte.

3.5 Vorhersagematrix

Wesentlich interessanter als ein Pattern ist allerdings eine quantitative Vorhersagematrix. Statt nur bestimmte Aminosäuren an den entsprechenden Positionen zuzulassen oder zu verbieten und damit nur eine ja-nein-Entscheidung zu fällen, erlaubt eine Vorhersagematrix das *ranking* von Peptiden, wobei der Einfluß einer „ungünstigen“ Aminosäure an einer Position durch eine „besonders günstige“ an einer anderen Position wieder aufgehoben werden kann.

Mit *Position specific scoring matrices* können bindende Peptide selbst dann erkannt werden, wenn keine Sequenzhomologie zwischen bekannten Bindern und dem Targetpeptid vorliegt. Die Methode wurde bereits zur Vorhersage von MHC-bindenden Peptiden [146] und Sulfatierung von Tyrosylresten [147] erfolgreich eingesetzt.

Die Methode setzt voraus, daß die Interaktion zwischen dem Peptid und Pex19 spezifisch von sogenannten *Ankerpositionen* abhängt, also keine unspezifische Bindung an z. B. hydrophobe Bereiche vorliegt. Die Analyse der Substitutionsmatrix (Abb. 3.4) zeigt, daß ein Fenster von neun Aminosäuren Länge offensichtlich für eine spezifische Bindung erforderlich ist – für jede der neun Positionen gelten spezifische Präferenzen für tolerierte Aminosäuren.

Der *score* eines Peptides durch Anwendung der Matrix ergibt sich dabei durch Potenzierung der Eulerschen Zahl e mit der Summe der Matrixeinträge der Aminosäuren über die Länge des Peptides

$$s = \exp\left(\sum_{\text{pos}} m_{aa,\text{pos}}\right)$$

Zunächst wurde unter Verwendung eines nichtlinearen Optimierungsalgorithmus' (FrontLine Systems, Solver DLL V 3.5, Incline Villa, NV; 1999) eine Matrix generiert, die die gemessenen Werte der Substitutionsanalyse (Abb. 3.4) optimal widerspiegelt. Dazu wurden die Matrixeinträge an neun Aminosäurepositionen derart optimiert, daß die Summe der Abstandsquadrate zwischen berechnetem und gemessenem Wert minimal wird.

Mit Hilfe dieser Matrix wurde die Sequenz des peroxisomalen Membranproteins Pex11 nach Pex19-Bindestellen durchsucht. Während die Anwendung des Patterns zwei Peptide als mögliche Binder ergibt (KVLRLLLQYLARFLAV an Position 26 und YVALSGVVTSILGMQ an Position 215), ergibt die Anwendung der Vorhersagematrix als besten Treffer das erste Peptid mit einem *score* von 12920. Das zweite durch den *Pattern match* gefundene Peptid erhält durch die Matrix-Methode einen *score* von 3590. Im Gegensatz zur Patternsuche, die zwischen den Treffern nicht unterscheidet, liegt bei Anwendung der Matrix eine Wichtung vor.

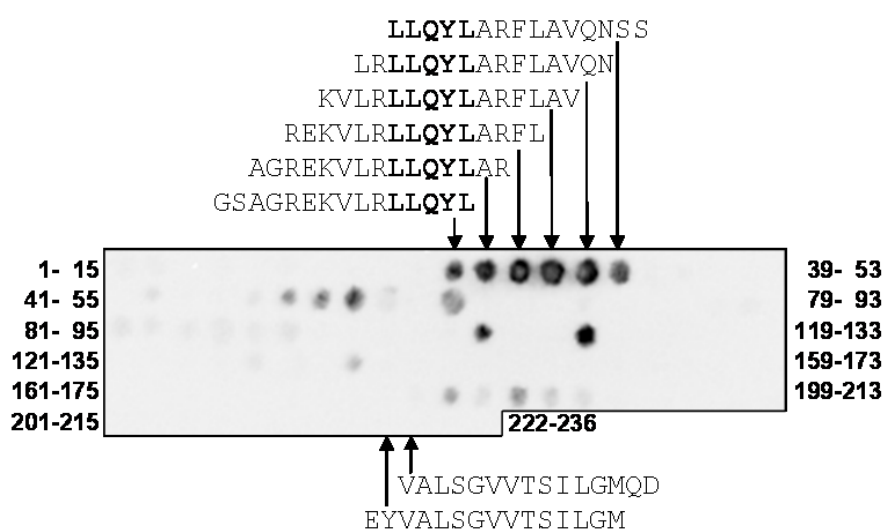


Abbildung 3.6: Peptidscan von Pex11. Mit einem Versatz von zwei Aminosäuren wurden Peptide der Länge 15 Aminosäuren synthetisiert, die das gesamte Protein abdecken. Man erkennt deutlich einen Bindebereich.

Der *Peptidscan* durch Pex11 (Abb. 3.6) zeigt deutlich einen Bindebereich, der das erste Peptid umfasst. Peptide im Bereich des zweiten *Pattern match* zeigen keine Bindung. Auffällig ist hier, daß der bindende Bereich von Pex11 sechs Spots umfasst, die sich nur in fünf Aminosäuren überlappen.

Um die Spezifität der Pex19-Bindestelle weiter zu untersuchen, wurde analog zur Mutationsanalyse des Peptides aus Pex13 auch das bindende Peptid aus Pex11 untersucht (Abb. 3.7).

Besonders deutlich tritt auch hier die hohe Restriktivität der Aminosäure an Position neun auf, wo eine verzweigtkettige aliphatische Aminosäure bevorzugt wird. Ähnliche Präferenzen treten auch an Positionen fünf und sechs auf. Wie bei dem Peptid aus Pex13 sind auch hier die ersten und letzten Aminosäuren nicht kritisch für eine Bindung an Pex19.

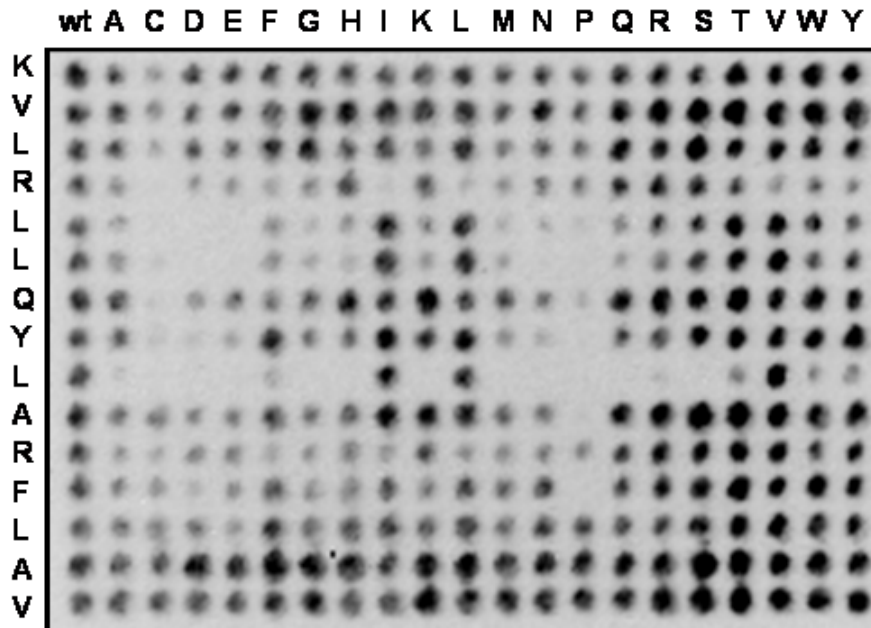


Abbildung 3.7: Mutationsanalyse der *Pex19*-Bindestelle aus *Pex11*. Die an den jeweiligen Positionen tolerierten Aminosäuren ähneln den Präferenzen aus dem *Pex13*-Peptid.

Wie kann die Information der neuen Experimente in die Vorhersagematrix übernommen werden? Da die beiden Mutationsanalysen und die experimentelle Testung der Pattern-Hits unabhängige Experimente darstellen, sind die auf den jeweiligen Membranen gemessenen Intensitäten nicht miteinander vergleichbar.

Soll durch Minimierung eine Vorhersagematrix gefunden werden, die Kriterien aus allen drei Experimenten berücksichtigt, so muß für jedes der drei Experimente ein frei skalierbarer Faktor eingebunden werden, mit dem die Ergebnisse der Vorhersage multipliziert werden, um den unterschiedlichen Intensitäten der Meßwerte auf den Membranen Rechnung zu tragen.

In diesem Fall wurde der Faktor f für jede Membran so berechnet, daß die Summe der quadratischen Abweichungen für die jeweilige Membran minimal wird. Bei Vorhersagewerten p_i und Meßwerten m_i ergibt sich die Summe der quadratischen Abstände über alle Peptide zu

$$rmsd = \sum_{\text{pep}} (f \cdot p_i - m_i)^2$$

mit der Ableitung

$$\frac{\delta_{rmsd}}{\delta_f} = 2(f \cdot \sum_{\text{pep}} p_i^2 - \sum_{\text{pep}} m_i p_i).$$

Setzt man die erste Ableitung Null, ergibt sich der Faktor f zu

$$f = \frac{\sum_{\text{pep}} p_i m_i}{\sum_{\text{pep}} p_i^2}.$$

Einsetzen des optimalen Faktors in die Gleichung für $rmsd$ ergibt

$$\begin{aligned} rmsd &= \sum_{\text{pep}} (f \cdot p_i - m_i)^2 \\ &= f^2 \sum_{\text{pep}} p_i^2 - 2f \sum_{\text{pep}} p_i m_i + \sum_{\text{pep}} m_i^2 \\ &= \left(\frac{\sum_{\text{pep}} p_i m_i}{\sum_{\text{pep}} p_i^2} \right)^2 \sum_{\text{pep}} p_i^2 - 2 \frac{\sum_{\text{pep}} p_i m_i}{\sum_{\text{pep}} p_i^2} \sum_{\text{pep}} p_i m_i + \sum_{\text{pep}} m_i^2 \\ &= - \frac{\left(\sum_{\text{pep}} p_i m_i \right)^2}{\sum_{\text{pep}} p_i^2} + \sum_{\text{pep}} m_i^2 \end{aligned}$$

Mittels dieser Gleichung läßt sich direkt (*on the fly*) der $rmsd$ unter Verwendung des optimalen Faktors f für jede Einzelmatrix berechnen.

Um möglichst viele Informationen aus Experimenten in die Vorhersagematrix einzubeziehen, wurde mittels des nichtlinearen Optimierungsprogrammes eine Matrix generiert, die mit möglichst geringem $rmsd$ aus den Sequenzen der Peptide die Meßwerte der Substitutionsmatrizen für die Peptide aus Pex13 (Abb. 3.4) und Pex11 (Abb. 3.7) sowie die Meßwerte der *Pattern hits* aus Abb. 3.5) generiert. Die optimierte Matrix ist in Tab. 3.3 gezeigt.

Die Matrix enthält nur Werte für die mittleren neun Positionen, die schon in der Mutationsanalyse des Pex13-Peptides hervortraten (Abb. 3.4). Matrixeinträge zu jeder Position wurden durch Subtraktion des Mittelwertes der jeweiligen Matrixeinträge auf den Mittelwert Null normiert. Es fällt auf, daß die neunte Position offensichtlich die stärkste Stringenz aufweist. Hydrophobe Aminosäuren an dieser Position begünstigen die Bindung an Pex19, polare Aminosäuren sind ungünstig.

Weitere stringente Positionen sind an Stelle fünf und zwölf zu finden. Insbesondere an der fünften Position sind die Aminosäurepräferenzen der neunten Position sehr ähnlich – hydrophobe Aminosäuren wirken sich günstig

Tabelle 3.3: Optimierte Matrix zur Vorhersage der Bindeaffinität von Peptiden zu Pex19. Werte kleiner als -1 und größer als 1 sind hervorgehoben.

	5	6	7	8	9	10	11	12	13
A	0.253	-0.193	0.437	-0.490	2.637	0.187	0.400	1.020	0.231
C	0.267	-0.628	<u>-1.308</u>	0.733	2.429	0.092	0.003	0.833	0.182
D	0.265	<u>-1.068</u>	-0.924	-0.510	<u>-1.160</u>	-0.295	0.029	-0.983	-0.684
E	-0.181	-0.386	0.064	-0.046	<u>-1.755</u>	<u>-1.857</u>	0.089	-0.056	-0.490
F	1.582	0.992	-0.601	0.839	2.072	0.608	0.388	1.184	0.144
G	-0.980	-0.317	-0.660	-0.762	<u>-1.796</u>	-0.064	-0.585	-0.951	-0.395
H	-0.762	0.022	0.490	-0.248	<u>-2.261</u>	-0.114	0.148	0.309	-0.263
I	1.178	0.686	0.272	0.936	2.870	0.581	0.013	1.102	0.717
K	<u>-2.758</u>	0.521	0.557	-0.642	<u>-2.226</u>	0.441	0.271	<u>-1.483</u>	-0.293
L	1.657	0.749	0.185	1.201	2.927	0.693	0.615	1.473	0.900
M	0.243	0.499	-0.221	-0.025	-0.012	-0.179	0.083	-0.011	-0.497
N	-0.453	<u>-2.062</u>	-0.521	-0.662	<u>-1.504</u>	<u>-1.218</u>	<u>-1.135</u>	-0.166	-0.283
P	<u>-1.919</u>	-0.590	-0.246	<u>-3.189</u>	<u>-1.771</u>	<u>-1.176</u>	-0.397	<u>-3.798</u>	<u>-1.060</u>
Q	-0.841	0.340	0.066	0.574	<u>-2.149</u>	0.259	0.420	0.408	0.922
R	-0.470	0.648	0.695	-0.428	<u>-1.471</u>	0.263	0.362	0.101	1.011
S	-0.258	-0.357	0.822	0.115	-0.997	0.179	0.179	0.253	0.099
T	1.284	0.067	0.290	0.239	-0.026	0.174	-0.227	0.126	-0.091
V	0.977	0.777	0.184	1.263	2.767	0.778	-0.122	0.803	0.695
W	0.754	0.372	0.426	0.802	0.976	0.612	0.113	1.154	-0.304
Y	0.163	-0.072	-0.005	0.299	0.448	0.035	-0.647	<u>-1.317</u>	-0.541

aus (insbesondere die verzweigtkettigen aliphatischen), polare Aminosäuren sind ungünstig.

Die Ähnlichkeit der Aminosäurepräferenzen an den Positionen fünf, neun und zwölf läßt auf eine helikale Struktur der Bindestelle schließen – in einer Helix liegen diese Positionen jeweils auf der gleichen Seite. Auch der an jeder Position auftretende negative Effekt des Prolins spricht für eine helikale Struktur, da diese Aminosäure Helices destabilisiert. Auch der typische *Helixbrecher* Glycin wirkt sich an jeder Position negativ aus.

Der prediktive Wert der Vorhersagematrix kann nun durch die Untersuchung weiterer Proteine festgestellt werden. Zu diesem Zweck wurden die Proteine Pex12 (Abb. 3.8), Pex25 (Abb. 3.9) und Pex27 (Abb. 3.10) abschnittsweise in 15er Peptides synthetisiert, die das gesamte Protein abdecken.

In allen drei Fällen kann der im Experiment gefundene Bindebereich durch die Matrix vorhergesagt werden. Es fällt auf, daß unter Verwendung der Vorhersagematrix oft mehrere Treffer hintereinander gefunden werden, typischerweise im Abstand von vier Aminosäuren. Der Grund dafür liegt in der Ähnlichkeit der Spalten fünf und neun in der Vorhersagematrix.

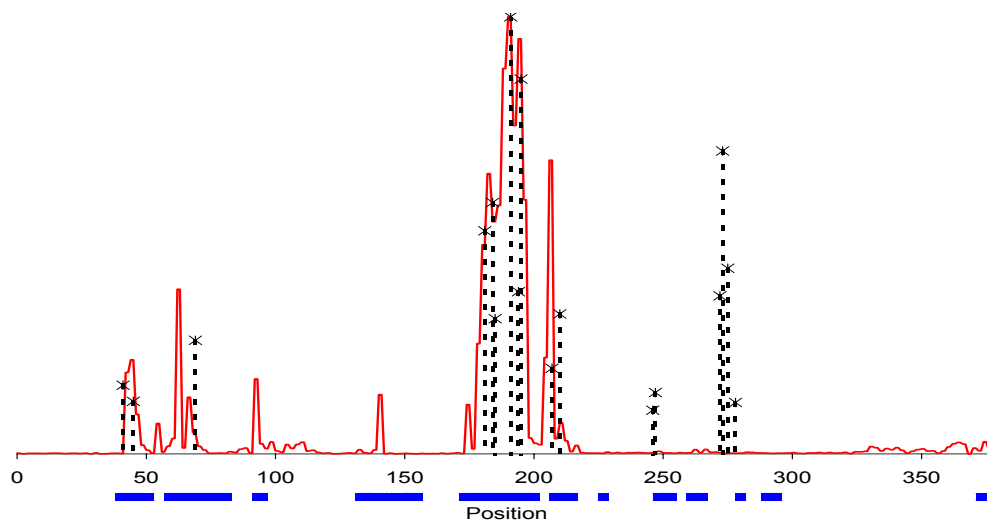


Abbildung 3.8: Peptidscan von Pex12 (rot) und vorhergesagte Bindeaffinitäten (gestrichelte Linien). Das Protein wurde mit einem Versatz von zwei Aminosäuren in 15er Peptiden synthetisiert. Die Größenordnung der frei skalierbaren Vorhersagewerte wurde den Meßwerten angepasst. Blau: vorhergesagt als helikaler Bereich nach NNPredict [148]

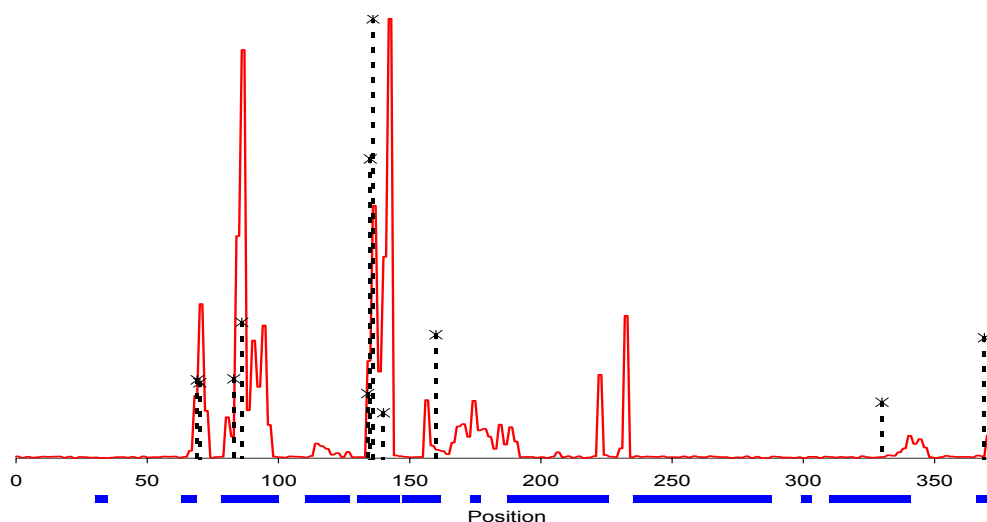


Abbildung 3.9: Peptidscan von Pex25. Vgl. Abb. 3.8

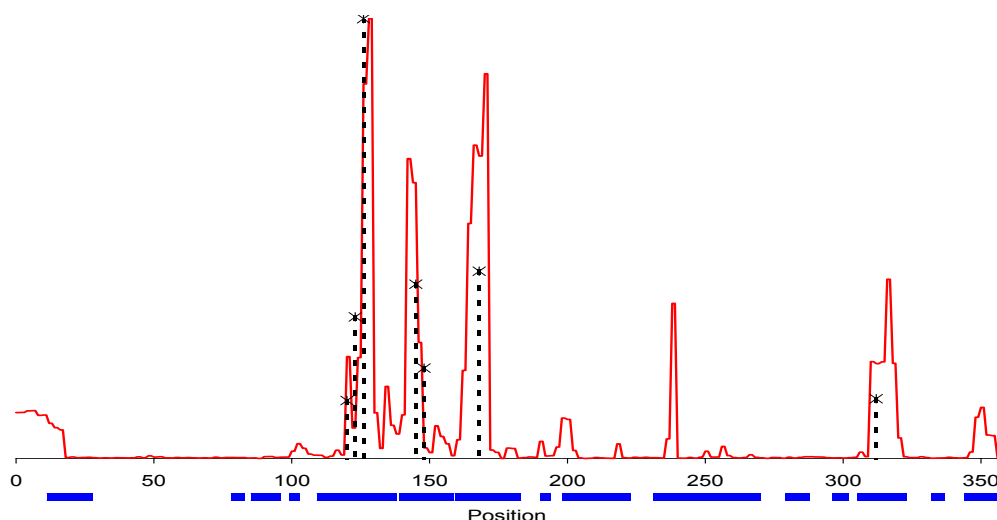


Abbildung 3.10: Peptidscan von Pex27. Vgl. Abb. 3.8

Weiterhin fällt auf, daß die angegebenen Bindebereiche oft $\approx \frac{1}{2}$ Äaere Abschnitte des Proteins umfassen. So erstreckt sich zum Beispiel der erste, große Bindebereich von Pex12 über zwanzig Positionen – die jeweils am Rand des Bindebereiches liegenden Peptide haben also keine Aminosäure gemein.

Die gefundene Länge der Bindebereiche und die Vielzahl der oft in Clustern auftretenden Treffer der Matrixvorhersage sprechen für eine Bindung von Pex19 an längere amphipathische helikale Bereiche.

Welche Positionen des Sequenzmotivs sind für die Bindung entscheidend? Um dieses Problem zu untersuchen, wurden unter Verwendung der gleichen Daten (Mutationsanalyse der Peptide aus Pex13 und Pex11 sowie Pattern-Hits) Matrizen mit sieben, acht, zehn und elf Spalten berechnet.

Bei der Verwendung von mehr Positionen besteht die Gefahr des *overfitting* – die erstellten Matrizen bilden zwar den Trainingsdatensatz besser ab, ihr prediktiver Wert ist aber kleiner. Die Auslassung von Sequenzpositionen führt zu einer Ausblendung von Informationen an den entsprechenden Stellen. Wenn diese Aminosäurepositionen für die Bindung wichtig sind, führt dies zu einer Verschlechterung der Vorhersage.

Der prediktive Wert der Matrizen kann nur visuell durch Vergleich mit *Peptidscans* ermittelt werden – eine berechnete Größe wie der *rmsd* trägt $\frac{1}{2}$ Äa nicht der Tatsache Rechnung, daß Peptide mit einer Verschiebung von wenigen Aminosäuren ebenfalls an Pex19 binden, jedoch nicht in das „Raster“ der Vorhersagematrix fallen.

Weder eine Verkleinerung noch eine Vergrößerung des untersuchten Se-

quenzfensters führte zu einer Verbesserung der Vorhersagen (nicht gezeigt). Die gewählten neun Positionen sind offenbar das bindungstragende Motiv.

Dem *overfitting* kann entgegengewirkt werden, indem bei der Minimierung des *rmsd* ein Strafterm für Abweichungen der Matrixeinträge von Null eingeführt wird [146, 149]. Dies verhindert, daß einzelne Matrixeinträge $\frac{1}{2}$ Å auf schwacher Datengrundlage unangebracht erhöht werden. Die Wahl eines kleinen Strafparameters beeinflusst die Optimierung der Matrix nur wenig, bei Erhöhung des Strafparameters wird die erhaltene Matrix der Nullmatrix ähnlicher und damit weniger selektiv. Die Wahl verschiedener Strafparameter für Matrizen mit unterschiedlicher Anzahl an Positionen führte nicht zu einer Verbesserung der Vorhersageergebnisse (nicht gezeigt).

3.6 Permutationen

Unterscheiden sich die einzelnen Aminosäurepositionen in der Pex19-Bindestelle von Position zu Position, oder sind Nachbarpositionen gegeneinander austauschbar? Wenn dem so ist, dann spräche dies eher für eine unspezifische Bindung (hydrophobe Bereiche, Chaperon-Funktion von Pex19 [135]), im entgegengesetzten Fall wäre eine spezifische Interaktion bestimmter Positionen der Bindesequenz mit Pex19 wahrscheinlicher.

Um dieser Frage nachzugehen, wurden paarweise Austausche von Aminosäuren im Bindepeptid von Pex13 durchgeführt. Die Austausche wurden komplett durchgeführt, d. h. auch bei einem Austausch einer Aminosäure mit sich selbst wurde der entsprechende *spot* auf der Membran synthetisiert.

Abbildung 3.11 zeigt am Beispiel des Lysins an Position 10, daß ein Austausch durch die $\frac{1}{2}$ Åliche Aminosäure Arginin, nicht aber durch die benachbarte Aminosäure Leucin möglich ist. Dies deutet auf eine spezifische Interaktion des Peptids mit Pex19 hin, bei der jede Aminosäure des bindenden Peptides einen spezifischen Interaktionspartner in Pex19 besitzt. Die unterschiedlichen Meßwerte bei Peptiden gleicher Sequenz (K gegen K ausgetauscht) sind auf experimentelle Fehler der Methode zurückzuführen.

Die Austausche anderer Aminosäuren führen zu ähnlichen Ergebnissen – stets ist ein Austausch ähnlicher Aminosäuren möglich, ein Platztausch benachbarter Aminosäuren verringert die Affinität zu Pex19 stark. Die verminderte Bindungsfähigkeit zu Pex19 steht auch mit der postulierten helikalen Struktur des Peptids in Einklang – eine Verschiebung um eine Aminosäure führt zu einer stark veränderten Ausrichtung der Seitenkette.

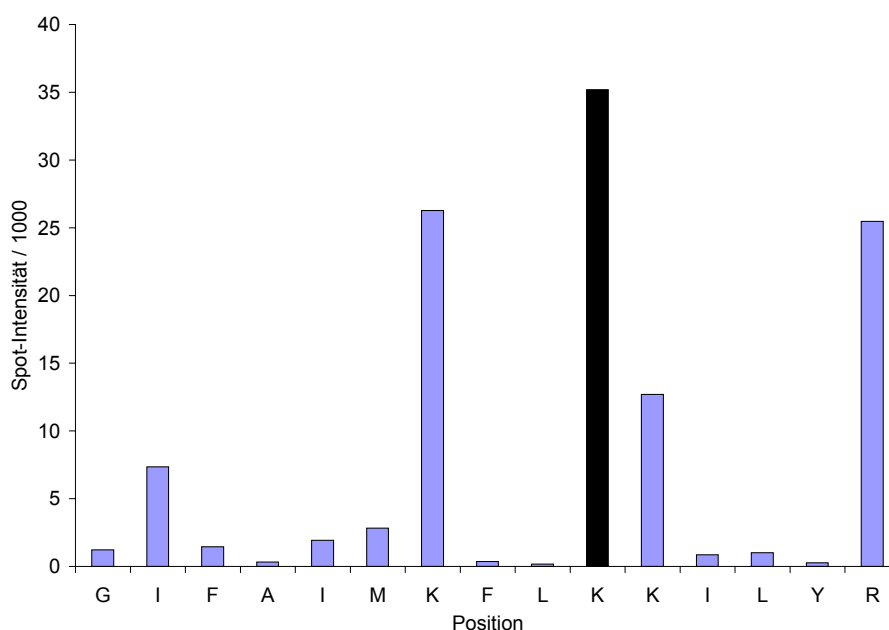


Abbildung 3.11: Permutation des Pex13-Bindepeptides. Beispielhaft ist hier der Austausch von Lysin an Position 10 mit allen anderen Aminosäuren des Peptids gezeigt.

3.7 Konservierung der Bindestelle

Die bisher geschilderten Experimente wurden mit Peroxinen der Hefe *S. cerevisiae* durchgeführt. Im weiteren Verlauf der Untersuchung soll nun geklärt werden, ob das Bindemotiv auch beim Menschen konserviert ist, d. h. ob die aus Hefedaten gewonnene Vorhersagematrix auch die Bindestellen der humanen Proteine an humanes Pex19 richtig erkennt.

Dazu wurden die bekannten humanen peroxisomalen Membranproteine mittels der Vorhersagematrix nach Bindestellen durchsucht (Tab. 3.4). Wie auch bei den Hefeproteinen, finden sich häufig in der Nachbarschaft von Peptiden mit hohem *score* weitere Peptide, die ebenfalls eine hohe Bewertung durch die Matrix erhalten. Die Sequenzen dieser randständigen Peptide wurden in der Tabelle mit berücksichtigt. Vielfach liegen die vorhergesagten Bindbereiche in Regionen, die für Targeting (Tab. 3.1) oder Bindung (Tab. 3.2) von Peroxinen von Bedeutung sind.

Um die Vorhersagen zu prüfen, wurden zu jedem Treffer neun 15 Aminosäure lange Peptide mit einem Versatz von je zwei Aminosäuren synthetisiert und auf die Bindung an humanes Pex19 getestet (Abb. 3.12).

Tabelle 3.4: Vorhergesagte Bindestellen humaner peroxisomaler Membranproteine. Angegeben sind jeweils die Positionen und Sequenzen der Peptide mit den höchsten scores innerhalb eines Proteins. Im Falle von breiteren Peaks ist die erweiterte Sequenz dargestellt. Helix: helikaler Anteil des core-Motivs nach NNPredict [148]

Protein	Position	Sequenz	Score	Helix
PEX2	147–161	KLGGFLNFIIFLQRG	7714	0,40
PEX2	199–213	lwhGFAEFLIFLLPLINV	7014	0,73
PEX2	52–66	FEPEVKACLWVFLWRftiy	3522	0,80
PEX10	238–252	syRLLGVISLLHLVLSMglq	6736	0,80
PEX10	157–171	qRRALLRAVFVLRQGL	4882	1,00
PEX11 α	173–187	eETEWLQSFLLLLFrs	13473	1,00
PEX11 β	236–250	GLCGLVSSILSILTliypwlr1	9095	0,20
PEX11 β	186–200	pqLALKLRLQVLLARVlrgh	6931	1,00
PEX11 γ	126–140	rwWTLSTTLWALSLLG	3714	0,73
PEX11 γ	141–155	VARSLWMLLKLRL	3024	0,87
PEX12	58–72	WFDEIFTLLDLLLQqhy	10723	1,00
PEX13	177–191	lkihftKVFSALVRTIRYLyrrlqrm1glrrgs	2437	1,00
PEX14	24–38	PREPLIATAVKFLQNs	6559	0,40
PEX16	119–133	vwgevgrwlvialiq1AKAVLRMLLLWFKA	6749	1,00
PEX26	276–290	aspsSLHFLYKLAQLFRWIrkaafs	8507	0,93
PEX26	253–267	fslpsKKSLLAALILCLLVVrfdpa	8439	0,73
PMP22	103–117	PLAGLRRLLLDRLVFa	6974	0,73
PMP22	18–32	PRRALAQYLLFLRLYp	4105	0,93
PMP24	127–141	iNMYLLSRVLFALSRL	6399	0,87
PMP24	155–169	PFPLLTAVVWGLVLWlfeyh	4384	0,53
PMP24	6–20	appQLRALLVVNALLRK	3550	1,00
PMP34	271–285	1EAKLLQTVLTAALMFlvyE	7875	1,00
PMP69	581–595	SLEKFHSLVLKLCGG	8221	0,67
PMP69	316–330	vsknafVCIYLLISCFTQLIDLstt	6342	0,07
PMP69	39–53	ssqNALMFTLLCLTLLEqf	4828	0,93
PMP69	20–34	rprld1qFLQRFLQILKVLFPswss	3624	0,73
PMP70	82–96	fckETGYLVLIAMVLVSR	10110	0,33
PMP70	59–73	VDKVFFSRLIQILKImvpr	7314	0,80
ALDP	68–82	NRVFLQRLLWLLRLLfprv1c	37393	0,87
ALDRP	80–94	VNADFFKQLELRKI	19872	0,80

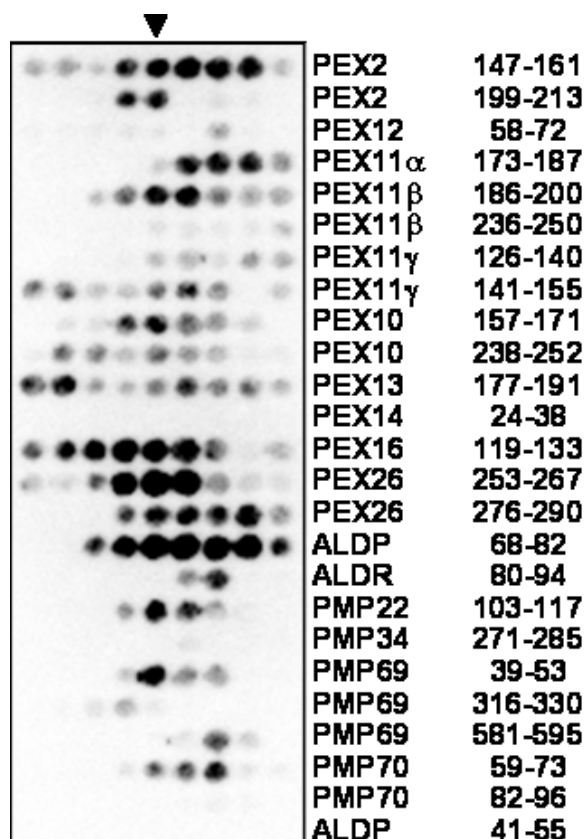


Abbildung 3.12: Interaktion vorhergesagter humaner Bindebereiche mit humanem Pex19. Neben dem zentralen Peptid aus Tab. 3.4 (Pfeil) wurden zu jeder Seite je vier Nachbarpeptide mit zwei Aminosäuren Versatz synthetisiert. Das letzte Peptid (ALDP₄₁₋₅₅) dient als Negativkontrolle.

In den meisten Fällen interagieren die vorhergesagten Bereiche mit Pex19. Die Konservierung der Bindestelle zwischen Hefe und Mensch konnte somit gezeigt werden. Durch *Yeast two hybrid* Technik und zelluläre Lokalisationsstudien konnte gezeigt werden, daß Bindebereiche von Mensch und Hefe jeweils wechselseitig auch mit dem jeweils anderen Pex19 interagieren [150]. Dies ist insofern erstaunlich, als die Sequenzidentität der beiden Proteine nur 20% beträgt.

Exemplarisch wurde das Protein Pex26 weiter untersucht, da bislang nichts über das Targeting-Signal dieses Proteins bekannt ist. Der Peptidscan (Abb. 3.13) des humanen Proteins zeigt sowohl in der Vorhersage als auch im Experiment zwei Bindestellen am C-Terminus des Proteins.

Durch *Yeast Two Hybrid* Experimente konnte gezeigt werden, daß ein C-terminales Fragment von Pex26, das die beiden Bindestellen umfasst, für

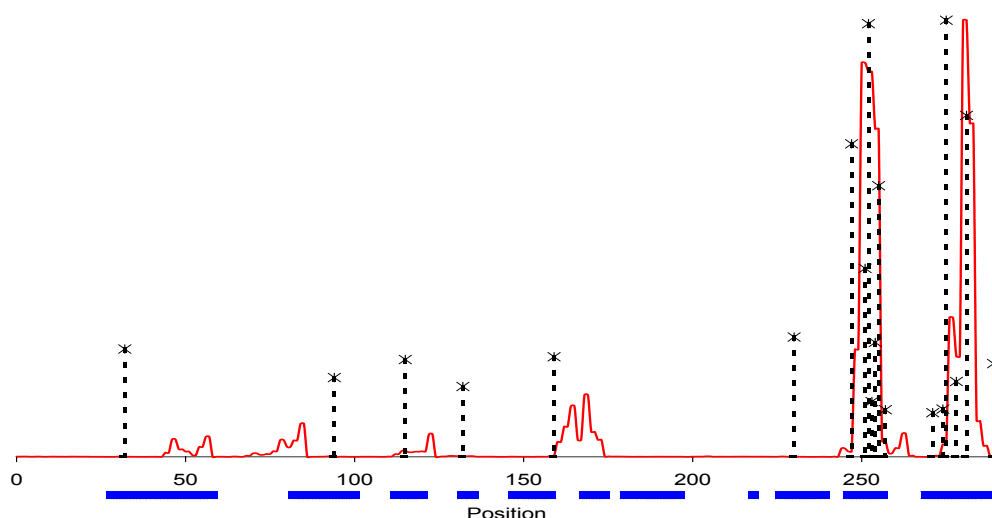


Abbildung 3.13: Peptidscan von Pex26. Vgl. Abb. 3.8

eine Interaktion mit Pex19 hinreichend ist. Auch die zweite Bindestelle allein ist für eine Interaktion hinreichend [150].

3.8 Diskussion

Ausgehend von einer Analyse von Peptiden aus dem peroxisomalen Membranprotein Pex13 konnte der Pex19-Bindebereich des Proteins aufgefunden und auf neun zentrale Aminosäuren eingegrenzt werden. Eine Längenanalyse zeigte, daß selbst Peptide aus elf Aminosäuren, die das zentrale Peptid enthalten, Pex19 binden. Dies ist ein Indiz für eine spezifische Bindung, im Gegensatz zur vorgeschlagenen Funktion von Pex19 als unspezifisches Chaperon [135].

Der für die Pex19-Bindung verantwortliche Bereich in Pex13 konnte auch durch die Analyse von Verkürzungskonstrukten aus Pex13 mittels *Yeast Two Hybrid* Technik nachgewiesen werden. Eine Bindung des entsprechenden Bereichs an Pex19 findet also offensichtlich auch unter *in vitro*-Bedingungen statt.

Die Mutationsanalyse des Bindebereichs aus Pex13 zeigt, daß hydrophobe Reste an verschiedenen Positionen eine Schlüsselrolle einnehmen. Ausgehend von den Informationen aus der Mutationsanalyse konnte ein Pattern erstellt werden, das die Suche nach Bindestellen in weiteren Proteinen ermöglichte.

Unter Verwendung experimenteller Daten aus Mutationsanalysen und

Peptiden, die dem Pattern entsprechen, konnte eine Vorhersagematrix zur Identifizierung von Pex19-Bindestellen in peroxisomalen Membranproteinen generiert werden. Die Matrix zeigt ähnliche Aminosäurepräferenzen an Positionen fünf, neun und zwölf, was auf eine amphipathische Helix hinweist. Die vorhergesagten Bindestellen in Proteinen der Bäckerhefe *S. cerevisiae* und im Menschen konnten experimentell validiert werden.

Auch durch Austausche von Aminosäuren innerhalb von bindenden Peptiden konnte die Spezifität der Bindung gezeigt werden: Ein Austausch ähnlicher Aminosäuren wird toleriert, der Austausch benachbarter Aminosäuren nicht. Dies ist ein weiteres Indiz für die helikale Struktur des Bindebereiches, da ein Versatz von einer Aminosäure zu einer veränderten Ausrichtung der Seitenketten führt. Diese Befunde sprechen für eine spezifische Bindung von Pex19 an Targetsignale und lassen eine Funktion von Pex19 als unspezifisches Chaperon [135] unwahrscheinlich erscheinen.

Bislang ist aber nur das Auffinden von bindenden Bereichen innerhalb eines Proteins möglich – das Durchsuchen kompletter Proteome nach Bindestellen liefert zu viele unspezifische Treffer. Unter den am höchsten bewerteten Bindebereichen in Proteinen findet sich kein höherer Anteil an peroxisomalen Membranproteinen als in Kontrollgruppen (nicht gezeigt).

Eine weitere Steigerung der Spezifität der Vorhersage ist daher wünschenswert. Es wäre möglich, zusätzlich zur Bewertung durch die Vorhersagematrix noch Ausschlußkriterien einzuführen, die die Anzahl gefundener Treffer weiter limitieren. Denkbar wäre zum Beispiel eine Einbeziehung anderer Vorhersagemethoden wie zum Beispiel des *Decision Tree Learning*.

Auch durch das Einbeziehen von multiplen Alignments könnte die Vorhersage noch weiter verbessert werden. Zu diesem Zwecke werden momentan multiple Alignments verschiedener Peroxine untersucht.

Kapitel 4

Lokale Sequenzähnlichkeiten

4.1 Vorhersagen aufgrund lokaler Sequenzumgebungen

Neben den im letzten Kapitel behandelten Signalsequenzen für das Protein Targeting lassen sich aus lokalen Sequenzumgebungen noch weitere Informationen ablesen. Ein weit verbreitetes Sequenzmotiv ist zum Beispiel Asn-X-Ser/Thr, dessen Asparaginrest häufig N-Glycosyliert wird [151]. Für die Effizienz der Glycosylierung ist die Aminosäure hinter dem Ser/Thr-Rest mit ausschlaggebend [152].

Ebenso zeigen die Phosphorylierungssites verschiedener Kinasen eine definierte Präferenz für die Sequenzumgebung von Serin-, Threonin- und Tyrosinresten, was die Vorhersage derartiger Sites in Proteinen erlaubt [153, 154]. Auch posttranslationale Modifikationen wie N-terminale Myristoylierung [155] und andere Lipidmodifikationen [156] sind durch spezifische Sequenzmotive codiert. Auch die Ubiquitinmarkierung von Proteinen, die zum proteasomalen Abbau führt, wird durch ein kurzes Sequenzmotiv (PEST) vermittelt [157].

Neben posttranslationalen Modifikationen läßt sich auch die Sekundärstruktur von Proteinen bereits mit einer Trefferquote von über 76% vorhersagen [158]. Bei der Sekundärstrukturvorhersage fließen allerdings neben lokalen Sequenzcharakteristika auch *long range*-Effekte und Informationen aus multiplen Alignments mit ein. Daß die bloße lokale Sequenzumgebung die Sekundärstruktur im Protein nicht allein determiniert, wird auch durch die bereits beschriebenen Chamäleon-Sequenzen (s. S. 18) deutlich.

Hier soll untersucht werden, inwieweit ein weiteres Charakteristikum der lokalen Faltung, die Konformation von Prolylbindungen in Proteinen, durch lokale Sequenzumgebungen bestimmt wird.

4.2 Bedeutung von *cis*-Prolylbindungen

Durch Mesomerieeffekte hat die Peptidbindung teilweise Doppelbindungscharakter, was dazu führt, daß der Torsionswinkel ω nur Werte um 0° und um 180° annimmt (Abb. 4.1).

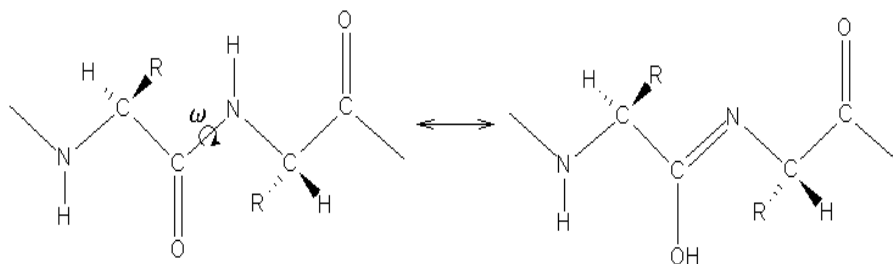


Abbildung 4.1: Durch Mesomeriestabilisierung (Delokalisierung des freien Elektronen-paars am Stickstoffatom) ist die Peptidbindung planar.

Die große Mehrheit der Peptidbindungen nimmt die *trans*-Konformation ($\omega=180^\circ$) ein. Nur 0,28% der Peptidbindungen finden sich in *cis*-Konformation, während im Gegensatz zu anderen Aminosäuren Prolin in über 5% der Fälle *cis*-Peptidbindungen eingeht [38]. Da die Peptidbindung in früheren Arbeiten grundsätzlich als *trans* angenommen wurde [37], sind *cis*-Prolylreste oft „übersehen“ worden. Heute weiß man, daß *cis*-Prolylreste wichtige Aufgaben in vielen Proteinen erfüllen.

So ist zum Beispiel ein konservierter *cis*-Prolylrest im inhibitorischen Loop des Bowman-Birk Proteaseinhibitors essentiell für dessen Funktion. Wird der entsprechende Prolylrest durch Alanin ausgetauscht, bildet sich eine *trans*-Peptidbindung an der entsprechenden Stelle, und das Protein ist inaktiv; Mutationen in der Umgebung des *cis*-Prolylrestes destabilisieren die *cis*-Konformation und führen zu einer Verringerung der Aktivität, da im Gleichgewicht nur ein Teil der Proteine in der *cis*-Konformation vorliegt. Es konnte gezeigt werden, daß Trypsin selektiv mit der *cis*-Form des Proteins interagiert [159].

Das „Umklappen“ der Prolylbindung kann auch als molekularer Schalter dienen – so zum Beispiel im Prothrombin, wo nach Calciumbindung eine *trans*-Prolylbindung in die *cis*-Konformation wechselt und so die Membranverankerung des Proteins ermöglicht wird [160]. Auch im Mannosebindenden Protein wird durch Calciumbindung eine Prolylbindung isomerisiert [161]. Eine spezielle Peptidylprolyl-*cis-trans*-Isomerase, Pin1, reguliert durch Konformationsänderung an Prolylbindungen eine Reihe von Substraten, unter anderem NIMA-Kinase [162], Cdc25, Myt1, Wee1, Plk1 und Cdc27 [163]

und verhindert dadurch den Eintritt der Zelle in die Mitose. Auch eine Interaktion mit RNA-Polymerase II [164] sowie Histon-Deacetylase [165] und dadurch eine direkte Regulation der Transkription konnte nachgewiesen werden. Scheinbar interagiert Pin1 auch mit phosphoryliertem Tau-Protein, das in der Alzheimer-Krankheit eine Rolle spielt [166, 167].

Für einige Proteine wie z. B. DsbA [168] konnte gezeigt werden, daß die Ausbildung einer *cis*-Peptidylbindung für die Stabilität des Proteins entscheidend ist – nach Mutation des entsprechenden Prolylrestes war die Stabilität der Proteine merklich vermindert. In Ribonuclease T1 [169], Aspartat Aminotransferase [170] und Aspartat Transcarbamoylase [171] hingegen bleibt die *cis*-Peptidylbindung auch erhalten, wenn der beteiligte Prolinrest zu Alanin mutiert wird.

Schon in frühen Arbeiten konnte gezeigt werden, daß die Ausbildung der *cis*-Prolylbindung eine Verzögerung im Faltungsvorgang von Proteinen wie Thioredoxin [172], Ribonuclease T1 [173] und Staphylokokken-Nuclease [174] hervorruft.

Daten zu Häufigkeit und Sequenzumgebungen von *cis*-Prolylresten existieren von mehreren Gruppen. Die beobachtete Häufigkeit von *cis*-Prolylresten divergierte aufgrund der kleinen Zahl von Proteinstrukturen in frühen Studien. Stewart *et al.* gaben 1990 einen Anteil von 6,5% *cis*-Prolylbindungen an [175], Frömmel und Preissner im gleichen Jahr 6% [39], während Weiss *et al.* 1998 5.21% der Prolylreste in *cis*-Konformation vorfanden.

Mehrere Gruppen untersuchten bereits die Sequenzumgebungen von *cis*-Prolylresten und versuchten, sequenzbasierte Regeln für das Auftreten eines *cis*-Prolylrestes zu finden [39, 176, 177]. Der generelle Befund war stets eine Häufung aromatischer Aminosäuren an der Position vor dem *cis*-Prolylrest.

Eine anderen Herangehensweise ist die Vorhersage mittels *Support Vector Machines*. Informationen über die zu beurteilenden Sequenzen werden dabei durch Anwendung einer mathematischen Funktion in einen mehrdimensionalen Raum (*feature space*) übertragen, in dem eine Trennung der beiden Klassen durch eine Hyperfläche möglich ist. Die an einem Lerndatensatz ermittelte Hyperfläche kann daraufhin zur Differenzierung der Proteine in Testdatensätzen verwendet werden. Die Methode wurde von Wang *et al.* zur Vorhersage von *cis*-Prolylresten verwendet [178].

Nicht nur die Häufigkeit, auch die Analyse der Sequenzumgebungen divergierte zwischen den verschiedenen Autoren: Stewart *et al.* [175] fanden Trp-Pro-Bindungen niemals in *cis*-Konformation, Tyr-Pro hingegen in 20% der Fälle; Pal und Mitarbeiter [176] fanden in mehr als 10% der *cis*-Prolylreste Tryptophan als Vorgänger des Prolins, während die Tyr-Pro-Bindung nur in 9,7% der Fälle die *cis*-Konformation einnahm.

Eine bedeutende Frage ist hier, ob nur die direkte Sequenzumgebung

des Prolylrestes oder komplexere weitreichende Wechselwirkungen die Ausbildung von *cis*-Prolylresten steuern. Untersuchungen an Peptiden deuten darauf hin, daß die Aminosäure vor Prolin den Gehalt an *cis*-Prolylresten qualitativ in derselben Weise beeinflußt wie in Proteinen [177].

4.3 Zielsetzung

Es soll untersucht werden, inwieweit *cis*-Prolylreste in homologen Proteinen konserviert sind. Unterschiede in der lokalen Sequenzumgebung von *cis*- und *trans*-Prolylresten sollen analysiert werden, auch im Hinblick darauf, ob die Ausbildung einer *cis*-Prolyl Bindung eher durch lokale oder globale Effekte vermittelt wird. Interessant ist auch die Frage, ob sich die Konformation der Peptidbindung aus lokalen Sequenzmustern vorhersagen läßt. Besonders für das *homology modelling* ist die Erkennung von *cis*-Prolylresten von großer Wichtigkeit.

4.4 *cis*-Prolylbindungen in bekannten Proteinstrukturen

4.4.1 Abhängigkeit von der Auflösung

Nichtredundante Teilmengen der Protein Data Bank [6] mit unterschiedlichen Schwellenwerten der gegenseitigen Sequenzidentitäten und unterschiedlichen Auflösungsgrenzen wurden vom PISCES-Server [179] bezogen.

Zunächst wurde die Abhängigkeit des Anteils an *cis*-Prolylresten von der Auflösung der Proteinstrukturen untersucht. Datensätze von Proteinen mit unter 25% paarweiser Sequenzidentität und unterschiedlichen Auflösungsgrenzen wurden auf ihren Gehalt an *cis*-Prolylresten überprüft (Abb. 4.2). Ein Torsionswinkel der CO–NH-Bindung (ω) von -45° bis 45° wurde dabei als *cis*-Prolylbindung gewertet, Winkel im Bereich von 135° bis 225° als *trans*. Größtenteils (ca. 97%) weichen die Torsionswinkel allerdings weniger als 10° von den idealen 0 bzw. 180° ab, 86% liegen sogar in einem Intervall der Breite $\pm 5^\circ$ um einen der beiden Winkel.

Wie bereits in vorherigen Studien [38, 175, 39] festgestellt, ist der Anteil an *cis*-Prolylbindungen in schlechter aufgelösten Strukturen kleiner als in gut aufgelösten Strukturen. Da die Peptidbindung in den meisten Fällen in der *trans*-Konformation vorliegt, wird bei Strukturen schlechter Auflösung oft nicht versucht, eine *cis*-Peptidylbindung zu modellieren, oder die Qualität der

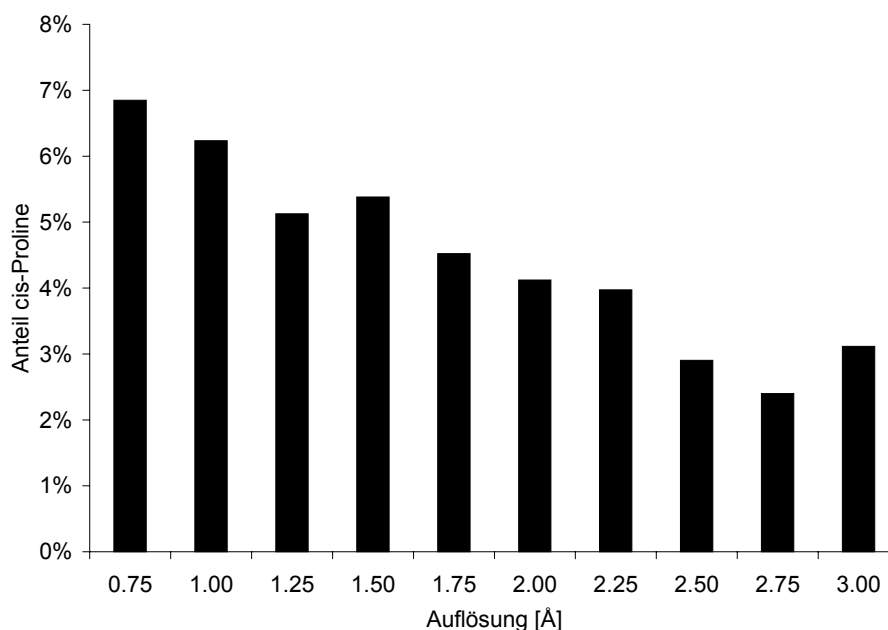


Abbildung 4.2: Anteil von *cis*-Prolylbindungen in Proteinen verschiedener Auflösungen. Bei schlechterer Auflösung sinkt der Anteil an *cis*-Prolylresten stark.

aufgezeichneten Reflexe läßt keine eindeutige Zuordnung der Konformation zu.

In Datensätzen mit kleiner paarweiser Sequenzidentität fanden wir nicht den publizierten hohen Anteil an *cis*-Prolylresten. In einem Datensatz mit unter 25% paarweiser Sequenzidentität und Auflösungen besser als 2 Å fanden Weiss *et al.* 6% der Proline in *cis*-Konformation [38], in dieser Arbeit betrug der Gesamtanteil bei diesen Proteinen nur 4,9%.

4.4.2 Abhängigkeit von paarweisen Sequenzidentitäten

Es wurde hingegen eine starke Abhängigkeit von der gewählten oberen Grenze der Sequenzidentitäten gefunden (Abb. 4.3). Der starke Anstieg des Anteils von *cis*-Prolylresten ist zu einem großen Teil auf die Anhäufung von Immunoproteinen zurückzuführen, die essentielle *cis*-Prolylreste besitzen: Ein Ausschluss sämtlicher Proteine mit den Schlüsselwörtern *FAB*, *antibody*, *MHC*, *histocompatibility* oder *HLA* führt zu einem Abfall von 5,72% auf 5,07% im Datensatz mit 95% Sequenzidentität. Proteine dieser Gruppe, die das Schlüsselwort „*antibody*“ enthalten, zeigen sogar einen Anteil an *cis*-Prolylresten von 18%.

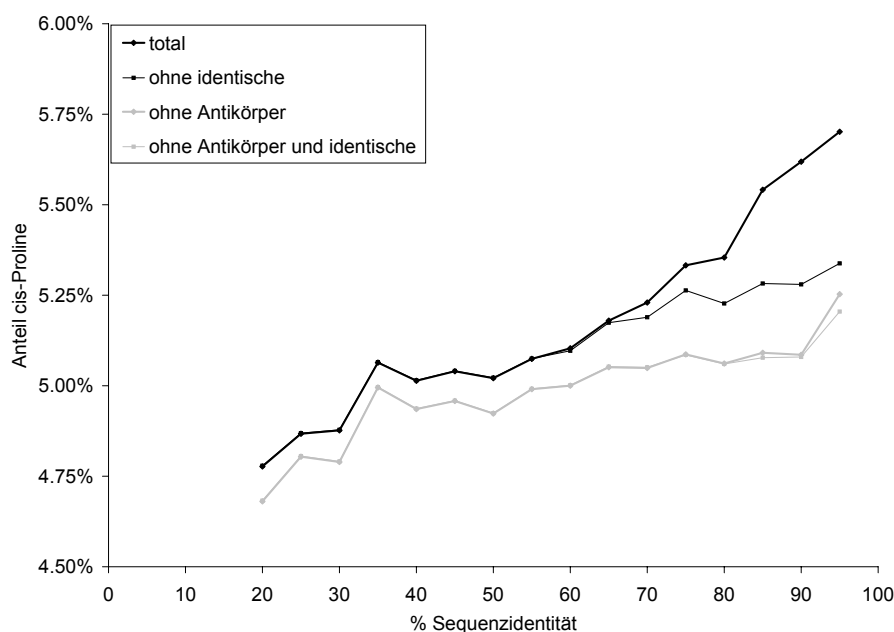


Abbildung 4.3: Abhängigkeit des Anteils an *cis*-Prolylresten vom Schwellenwert der Sequenzidentitäten innerhalb des Datensatzes. Ein Zulassen von größerer paarweiser Sequenzidentität führt zu einem Anstieg des Anteils. Dies ist zu einem großen Teil auf identische Peptidstücke in Proteinen (vgl. „ohne identische“) und zum Teil auf Immunoglobuline (vgl. „ohne Antikörper“) zurückzuführen.

Ein Ausschluß von Sequenzen, deren Aminosäuresequenz in einem Fenster von 21 Aminosäuren mit einer bereits im Datensatz enthaltenen Sequenz übereinstimmt, senkt den Anteil an *cis*-Prolylresten ebenfalls dramatisch ab (Abb. 4.3, „ohne identische“). Dies ist ein Indiz für die Häufung identischer oder sehr ähnlicher Proteine in der PDB, die einen hohen Anteil an *cis*-Prolylresten aufweisen.

4.4.3 Abhängigkeit von B-Faktoren

Ist der Anteil an *cis*-Prolylresten auch abhängig von den B-Faktoren der jeweiligen Aminosäuren? Der B-Faktor oder Temperaturfaktor ist ein Maß für die Flexibilität eines Atoms innerhalb des Kristalls und ist mit der mittleren Auslenkung u eines Atoms über die Beziehung $B = 8\pi^2 u^2$ verknüpft. Hohe B-Faktoren sind entweder auf hohe Flexibilität oder Fehler bei der Strukturaufklärung zurückzuführen.

Um die Abhängigkeit der Häufigkeit von *cis*-Prolylresten vom B-Faktor festzustellen, wurde die Konformation der Prolylbindungen in einem Daten-

satz von Proteinen mit einer Auflösung $\leq 2.0 \text{ \AA}$ bestimmt und dieser dann in 10 gleich große Gruppen nach B-Faktoren aufgeteilt (Abb. 4.4).

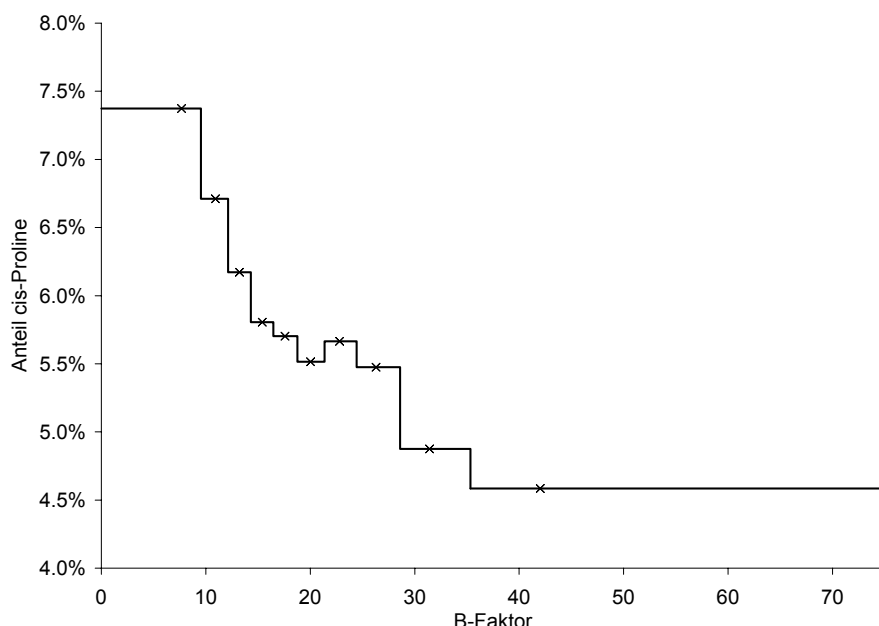


Abbildung 4.4: Prolylreste mit unterschiedlichen B-Faktoren des $C\alpha$ -Atoms zeigen deutlich unterschiedliche Anteile an *cis*-Peptidylbindungen. Gezeigt sind die jeweiligen Anteile nach Unterteilung des Datensatzes nach B-Faktoren in 10 gleich große Gruppen und die Mediane der jeweiligen Gruppen.

Es ist deutlich zu erkennen, daß Prolylreste mit kleineren Temperaturfaktoren im Durchschnitt einen größeren Anteil an *cis*-Peptidylbindungen aufweisen. Bei Prolinresten mit einem Temperaturfaktor unter 10 besitzen sogar 7,3% der Peptidylbindungen die *cis*-Konformation.

Die Abhängigkeit von B-Faktoren ist ein weiteres Indiz für die Tatsache, daß *cis*-Prolylreste in Proteinen bei der Strukturaufklärung oft unerkannt bleiben – eine irrtümliche Modellierung der entsprechenden Peptidylbindung in der *trans*-Konformation führt zu hohen B-Faktoren, da die Elektronendichte-Verteilung an dieser Stelle schlecht abgebildet wird.

4.4.4 Datensatz

Um zur Untersuchung nur Strukturen von guter Auflösung zu verwenden, Redundanz zu vermeiden und dennoch einen möglichst großen Datensatz zu erhalten, wurden Proteine mit weniger als 25% paarweiser Sequenzidentität, einer Auflösung von 2 \AA oder besser und einem R-Faktor von 0,25 oder

besser verwendet. Der R-Faktor ist ein Maß für die Übereinstimmung des theoretischen Beugungsmusters mit dem gemessenen Beugungsmuster:

$$R = \frac{\sum (|F_{obs}| - |F_{calc}|)}{\sum |F_{obs}|}.$$

Der Datensatz enthielt 1 729 Proteine mit 15 244 Prolinresten, die zu beiden Seiten mindestens 10 weitere Aminosäuren als Nachbarn besitzen. Von diesen waren 742 (4.87%) in *cis*-Konformation.

Zusätzlich zu Prolylresten in *cis*-Konformation wurden 118 nichtprolyl-*cis*-Aminosäuren gefunden (Tab. 4.1). Aus der Tabelle wird ersichtlich, daß die verzweigtkettigen Aminosäuren Isoleucin und Leucin seltener, Glutamin, Serin, Threonin und Tyrosin dagegen häufiger als erwartet in *cis*-Konformation auftreten.

Ein Grund für diese Häufung könnten die Wasserstoffbrücken-Donatoren der genannten Gruppen darstellen, die auf diese Weise eine Wechselwirkung mit der vorangehenden Aminosäure eingehen können, wenn diese durch die Ausbildung einer *cis*-Peptidylbindung in Nachbarschaft stehen.

4.5 Vorhersage mittels Propensities

4.5.1 Methode

Als *Propensity* einer Aminosäure an einer Stelle in der Sequenz bezeichnet man die Häufigkeit dieser Aminosäure an der entsprechenden Position in dem Teil des Datensatzes, der das interessierende Kriterium erfüllt, normiert auf die Häufigkeit in dem Teil des Datensatzes, der das Kriterium nicht erfüllt:

$$P_{Ala,3} = \frac{\frac{cis_{Ala,3}}{cis}}{\frac{trans_{Ala,3}}{trans}}$$

Tritt zum Beispiel eine Aminosäure an einer bestimmten Position innerhalb der Proteine mit *cis*-Prolylbindung doppelt so häufig auf wie in Proteinen mit *trans*-Prolylbindung, so ist die *Propensity* der Aminosäure an dieser Stelle zwei.

Gilt die Annahme, daß sich die Aminosäuren gegenseitig nicht beeinflussen, so läßt sich die *Propensity* eines Peptides durch Multiplizieren der *Propensities* der einzelnen Aminosäuren berechnen. Der erhaltene Wert ist dabei ein Maß dafür, wie wahrscheinlich es ist, daß das betreffende Peptid eine *cis*-Prolylbindung ausbildet:

Tabelle 4.1: Auftreten von non-Prolyl-cis-Aminosäuren im Datensatz mit 25% Sequenzidentitäts-Schwellenwert und erwartete Anzahlen aufgrund der Aminosäurehäufigkeiten.

Aminosäure	Anzahl	erwartet
A	10	9,4
C	1	1,9
D	8	6,6
E	6	6,4
F	7	5,9
G	14	13,2
H	3	3,4
I	2	7,5
K	6	6,0
L	2	9,9
M	0	2,2
N	10	5,8
Q	2	3,5
R	4	5,0
S	13	9,0
T	14	6,8
V	6	7,7
W	0	2,3
Y	10	5,4

$$p_{cis}^{mot} = \frac{P_{mot}}{P_{mot} + \frac{1-p_{cis}^{gen}}{p_{cis}^{gen}}}$$

Die Wahrscheinlichkeit, daß ein Sequenzmotiv zur Ausbildung einer *cis*-Peptidylbindung führt, hängt von der *Propensity* des Motivs P_{mot} und der relativen Häufigkeit von *cis*-Peptidylbindungen p_{cis}^{gen} ab.

Um die *Propensities* zu normalisieren, werden diese logarithmiert. Eine Aminosäure, die gleich häufig in *cis*- und *trans*- Peptiden vorkommt, erhält so den Wert Null. Die logarithmierten *Propensities* werden bei der Berechnung des Gesamtwertes für ein Peptid addiert, die Gesamt-*Propensity* eines Peptides ergibt sich als Exponentialwert dieser Summe.

Ein Schätzwert für die Qualität der Vorhersage mittels *Propensities* läßt sich gewinnen, indem man den Datensatz in zwei Klassen – einen Trainings-

und einen Testdatensatz einteilt. Die *Propensities* der einzelnen Positionen werden dabei anhand des Trainingssatzes ermittelt und die Konformation der Prolylbindung im Testdatensatz mit den erhaltenen *Propensities* vorhergesagt.

Die Güte der Vorhersage wurde mittels des AUC-Wertes [180] gemessen: Peptide wurden nach ihrem *score* (Gesamt-*Propensity*) sortiert und sukzessive zur Liste der als *cis*-Konformer vorhergesagten Peptide addiert. Nach jedem Schritt wurde die *specificity* (Anteil der korrekt zugeordneten *trans*-Peptide) gegen die *sensitivity* (Anteil der korrekt zugeordneten *cis*-Peptide) aufgetragen. Das Integral unter der erhaltenen Kurve ist ein Maß für die Güte der Vorhersage und entspricht der Wahrscheinlichkeit, daß ein zufällig ausgewähltes *cis*-Peptid einen höheren Score als ein zufällig ausgewähltes *trans*-Peptid besitzt.

4.5.2 Prediktiver Wert verschiedener Positionen

Welche Positionen der Sequenz sind für die Vorhersage wichtig? Zunächst sollen die Vorhersagegüten unter Verwendung verschiedener Sequenzpositionen verglichen werden (Abb. 4.5).

Die Größe des Trainings-Datensatzes wurde in Schrittweiten von 1% erhöht und jeweils 5 000 randomisierte Trainingssätze zum Lernen verwendet, die Abbildung zeigt die Vorhersagequalität (AUC-Wert) am verbleibenden Rest des Datensatzes.

Bei Einbeziehung von 10 Aminosäuren zu beiden Seiten des Prolylrestes in die Vorhersage wird bei hinreichend großem Trainingssatz ein AUC-Wert von 0.648 erreicht, mit nur zwei Positionen zu beiden Seiten des Prolylrestes erreicht der Wert 0.669. Mit der momentanen Größe des Datensatzes können offensichtlich nur Informationen über vier Aminosäurepositionen gewonnen werden, die Verwendung eines breiteren Sequenzfensters führt nicht zu einer Verbesserung der Vorhersage.

Selbst bei der Verwendung nur einer Aminosäureposition zu beiden Seiten des Prolylrestes wird die Vorhersage besser als mit 10 Positionen zu beiden Seiten – ein Indikator, daß weiter entfernt liegende Positionen das Ergebnis verrauschen.

Aus der Abbildung wird auch ersichtlich, daß der Trainingssatz für die Berechnung der *Propensities* für die beiden dem Prolinrest benachbarten Aminosäuren offensichtlich hinreichend groß ist, da die Vorhersagegüte bei größer werdendem Trainingssatz ein Plateau erreicht. Insbesondere die Vorhersage mit 10 Resten zu beiden Seiten des Prolins steigt aber bei größer werdendem Trainingssatz kontinuierlich und würde bei mehr verfügbaren Daten noch bessere Werte annehmen.

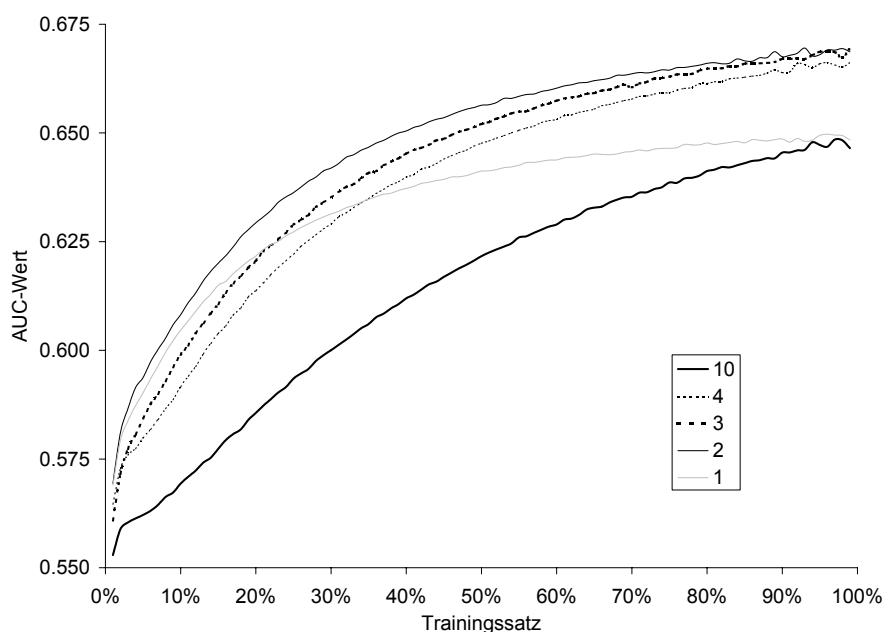


Abbildung 4.5: Vorhersagegüte für *cis*-Prolylreste unter Verwendung verschiedener Sequenzpositionen (1, 2, 3, 4 und 10 Reste zu beiden Seiten des Prolylrestes) in Abhängigkeit von der Größe des Trainings-Datensatzes. Die stabilste Vorhersage wird mit zwei Positionen zu beiden Seiten des Prolylrestes erreicht.

Abbildung 4.6 zeigt, daß auch für zwei Aminosäurepositionen auf jeder Seite des Prolylrestes die Lernmenge annähernd ausreichend ist: die Vorhersagegüte steigt bei größer werdendem Trainingssatz nicht mehr stark an, und auch die Vorhersagegüte des Trainingssatzes auf sich selbst erreicht ein Plateau.

Welche Aminosäureposition enthält die meiste Information? Um diese Frage zu klären, wurden unter Verwendung von nur je einer Aminosäureposition und einem Trainingssatz von 80% der verbleibende Testdatensatz untersucht (Abb. 4.7). Es wird deutlich ersichtlich, daß die Aminosäure vor dem Prolylrest (Position -1) die größte prediktive Aussagekraft besitzt. Dies deckt sich auch mit publizierten Sequenzauffälligkeiten an dieser Position [39, 176, 177].

Da offensichtlich die beiden dem Prolin am nächsten gelegenen Aminosäurepositionen die beste Vorhersage ermöglichen, wurde eine weitere Limitierung des Datensatzes unter Auslassung je einer weiteren Position unternommen (Abb. 4.8). Die Aminosäure zwei Positionen vor dem Prolin ist offensichtlich für die Vorhersage nicht von tragender Bedeutung. Bei der Verwendung großer Trainingssätze ist kein Unterschied gegenüber der Verwen-

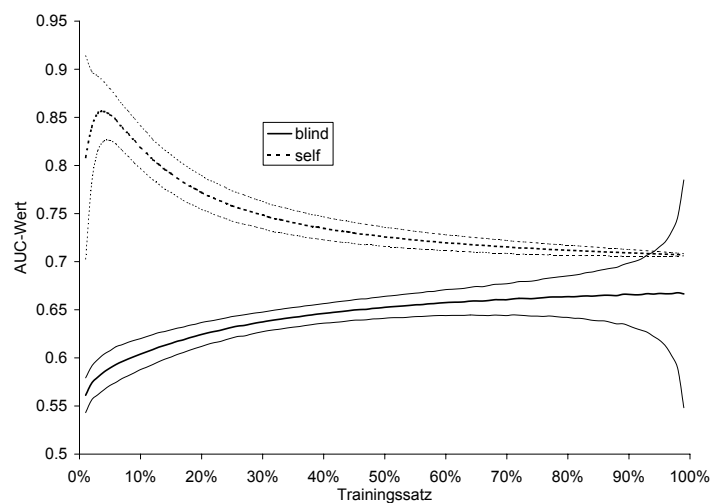


Abbildung 4.6: Vorhersagegüte des Trainingssatzes selbst (self) und des Testdatensatzes mit Standardabweichungen bei Verwendung von zwei Aminosäurepositionen zu beiden Seiten des Prolinrestes. Die Güte der Blindvorhersage nähert sich bei größerem Trainingssatz der Selbstvorhersage an.

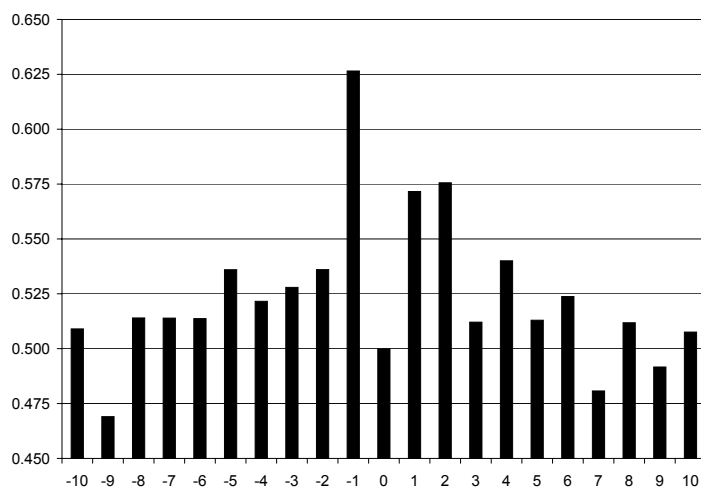


Abbildung 4.7: Prediktive Aussagekraft einzelner Aminosäurepositionen. Gezeigt ist der AUC-Wert der Vorhersage mit 80% Trainingssatz unter Verwendung nur jeweils einer Aminosäureposition.

dung von zwei Positionen auf jeder Seite erkennbar, bei kleineren Datensätzen ist die Vorhersage mit nur einer führenden und zwei dem Prolin folgenden Positionen allerdings besser.

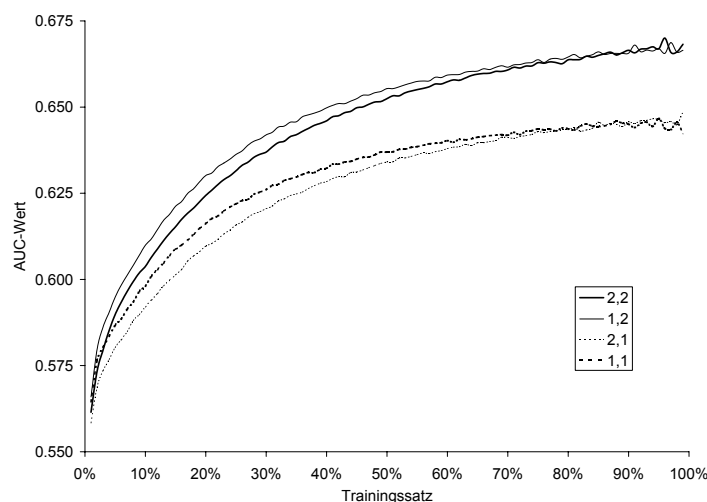


Abbildung 4.8: AUC-Werte bei Vorhersagen auf Grundlage asymmetrischer Aminosäurefenster. Angegeben sind jeweils die zur Vorhersage benutzten Aminosäuren vor und hinter dem Prolylrest. Bei der Verwendung der Position vor dem Prolylrest und zwei nachfolgenden Positionen (1,2) sind bei kleineren Trainingsätzen bessere Ergebnisse möglich als unter Verwendung von je zwei Positionen zu beiden Seiten des Prolylrestes.

In Tabelle 4.2 sind die Logarithmen der errechneten *Propensities* für zwei dem Prolin benachbarte Positionen unter Einbeziehung des gesamten Datensatzes gezeigt.

Die Tabelle zeigt die bereits bekannte [39, 176, 177, 181] Häufung aromatischer Aminosäuren und eine verringerte Anzahl an verzweigten aliphatischen Aminosäuren vor *cis*-Prolylresten. Desweiteren fällt auf, daß Cysteinreste im Abstand von zwei Aminosäuren vom Prolylrest ebenfalls die Ausbildung der *cis*-Prolylbindung begünstigen, ebenso wie ein weiterer Prolylrest zwei Positionen hinter dem Prolin.

4.6 Konservierung von Prolylresten

Eine weitere interessante Frage betrifft den Konservierungsgrad von Prolylresten im Vergleich zum Gesamtprotein. Da *cis*-Prolylreste oft wichtige Funktionen übernehmen (s. Abschnitt 4.2), steht zu erwarten, daß Prolylreste höher konserviert sind als andere Aminosäuren. Um diese Frage zu

Tabelle 4.2: Logarithmen der Propensities einzelner Aminosäuren in den zwei dem Prolin benachbarten Positionen. Werte über 0.5 und unter -0.5 sind hervorgehoben.

Aminosäure	-2	-1	1	2
A	0.03	0.26	0.00	<u>-0.52</u>
C	0.73	-0.27	0.44	0.54
D	-0.18	<u>-0.54</u>	-0.46	0.05
E	-0.22	0.33	<u>-0.84</u>	-0.49
F	-0.10	0.25	0.61	-0.06
G	-0.06	0.61	-0.02	-0.08
H	-0.05	-0.43	0.47	0.06
I	-0.31	<u>-1.05</u>	-0.10	-0.11
K	0.12	0.01	-0.25	0.05
L	-0.21	<u>-0.75</u>	0.18	-0.38
M	0.22	-0.13	0.02	-0.44
N	0.06	0.01	-0.08	0.37
P	0.27	0.49	0.30	0.83
Q	0.39	0.18	-0.15	-0.08
R	-0.38	-0.22	-0.26	0.12
S	0.33	0.07	0.07	-0.30
T	0.13	-0.29	0.12	0.27
V	-0.20	<u>-0.68</u>	-0.29	-0.04
W	-0.42	1.15	-0.03	-0.01
Y	0.04	0.84	0.70	0.13

untersuchen, wurden multiple Alignments der 1729 Proteine aus der HSSP-Datenbank [182] verglichen. In Abhängigkeit vom Konservierungsgrad der Gesamtproteine untereinander zeigt Abb. 4.9 die Konservierung der Prolyl- und Nachbarreste.

Es fällt auf, daß Prolylreste stärker konserviert sind als die jeweiligen Gesamtproteine. *Cis*-Prolylreste übertreffen in ihrer Konservierung noch die *trans*-Prolylreste. Dies ist ein weiterer Hinweis darauf, daß Prolylreste – und insbesondere *cis*-Prolylreste – wichtige Aufgaben in Proteinen übernehmen. Der Konservierungsgrad der benachbarten Aminosäuren unterscheidet sich dagegen nicht vom Konservierungsgrad der Proteine insgesamt.

In den Fällen, in denen die Prolylreste nicht konserviert sind, wurde die Austauschfrequenz durch unterschiedliche Aminosäuren untersucht. Gibt es Aminosäuren, die die kritischen Aufgaben von *cis*-Prolylresten besser erfüllen und daher diese häufiger ersetzen als andere?

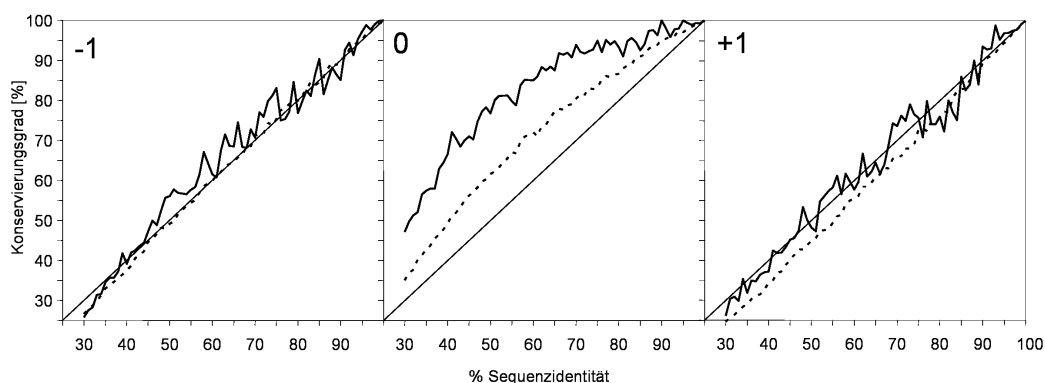


Abbildung 4.9: Grad der Konservierung von Aminosäuren vor (-1) und hinter (+1) Prolylresten sowie Prolylresten (0) im Vergleich zur Konservierung des Gesamtproteins. durchgezogene Linie: *cis*-Prolylreste bzw. benachbarte Aminosäuren; gestrichelte Linie: *trans*-Prolylreste bzw. benachbarte Aminosäuren. Die Prolylreste sind erheblich stärker konserviert als der Rest des Proteins.

Abbildung 4.10 zeigt Unterschiede in den Austauschfrequenzen von *cis*- und *trans*-Prolylresten ($p < 0.001$). Aminosäuren vor *cis*-Prolylresten werden offensichtlich eher durch Glycin und seltener durch verzweigtkettige aliphatische Aminosäuren ausgetauscht als Aminosäuren vor *trans*-Prolylresten. Dies kann auf die größere Flexibilität von Glycinresten zurückzuführen sein.

Auch *cis*-Prolylreste selber werden häufiger durch Glycin ausgetauscht als *trans*-Prolylreste. Die flexible Aminosäure ist offensichtlich besser geeignet, die Konformation des *cis*-Prolylrestes nachzuahmen.

Die Häufigkeiten der Austausche vor und hinter dem *cis*-Prolylrest ähneln qualitativ den in Tab. 4.2 aufgeführten *Propensities* – verzweigtkettige Aminosäuren sind seltener vor *cis*-Prolylresten zu finden, Glutamat seltener an der darauffolgenden Position.

4.7 Lokale Sequenzvergleiche

Wie die vergleichsweise geringen Vorhersagegüten nahelegen, sind die *Propensities* der benachbarten Aminosäuren allein kein guter Indikator für die Konformation von Prolylresten. Daher wurde im Folgenden untersucht, inwieweit Ähnlichkeiten in der lokalen Sequenzumgebung von Prolylresten Rückschlüsse auf die Konformation der Prolylbindung erlauben. Symmetrische Fenster steigender Größe um jeden *cis*-Prolylrest aus dem Datensatz wurden mit entsprechenden Fenstern um alle anderen Prolylreste in der PDB verglichen.

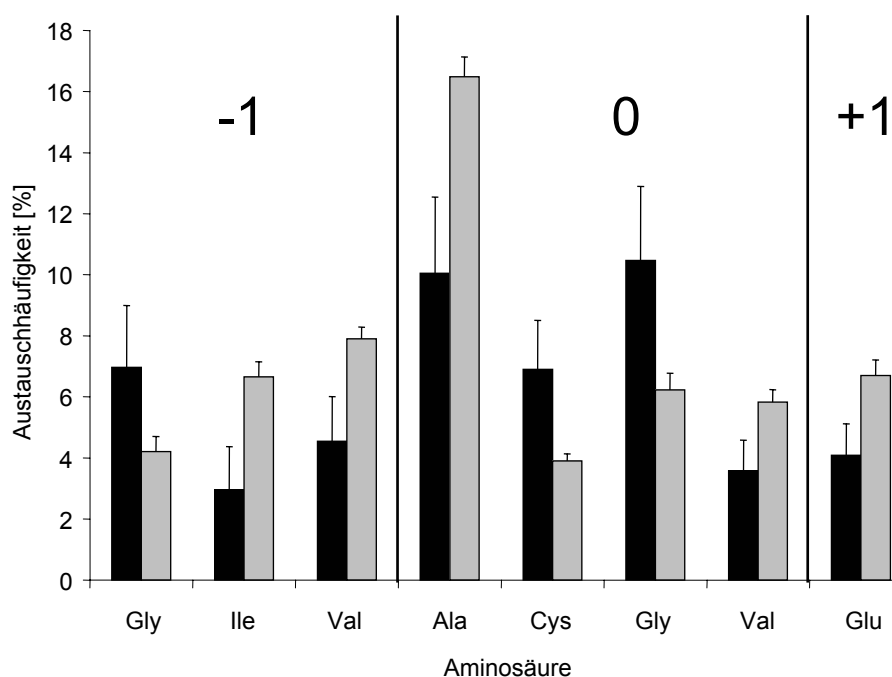


Abbildung 4.10: Austauschhäufigkeiten von *cis* (schwarze Balken) - und *trans* (graue Balken) -Prolylresten und benachbarten Aminosäuren.

Abb. 4.11 zeigt die Anteile an Prolylresten mit unterschiedlichen lokalen Sequenzähnlichkeiten zu *cis*-Prolylresten, die ebenfalls die *cis*-Konformation einnehmen. Ist eine lokale Sequenzidentität zu einem *cis*-Prolylrest ein Indikator für das Auftreten der *cis*-Konformation?

Die Abbildung zeigt, daß bei wachsender Fenstergröße um den interessierenden *cis*-Prolylrest selbst kleinere Sequenzidentitäten eine große Aussagekraft besitzen. Betrachtet man die Sequenzidentität über das gesamte Protein, sind selbst kleine Werte von 30% noch ein guter Indikator für die Konservierung des *cis*-Prolylrestes. Bei kleineren Fenstergrößen ist ein höherer Grad an Sequenzidentität notwendig, um auf die Konservierung des *cis*-Prolylrestes zu schließen.

4.8 Strukturelle Alignments

In der HSSP-Datenbank werden nur Alignments mit einer Sequenzidentität von >30% hinterlegt. Um auch Sequenzidentitäten von unter 30% zu untersuchen, wurden beispielhaft die in Tab. 4.3 aufgeführten Proteinfamilien aus der SCOP-Datenbank [183] mit dem CE-Algorithmus [50] überlagert (Abb. 4.12).

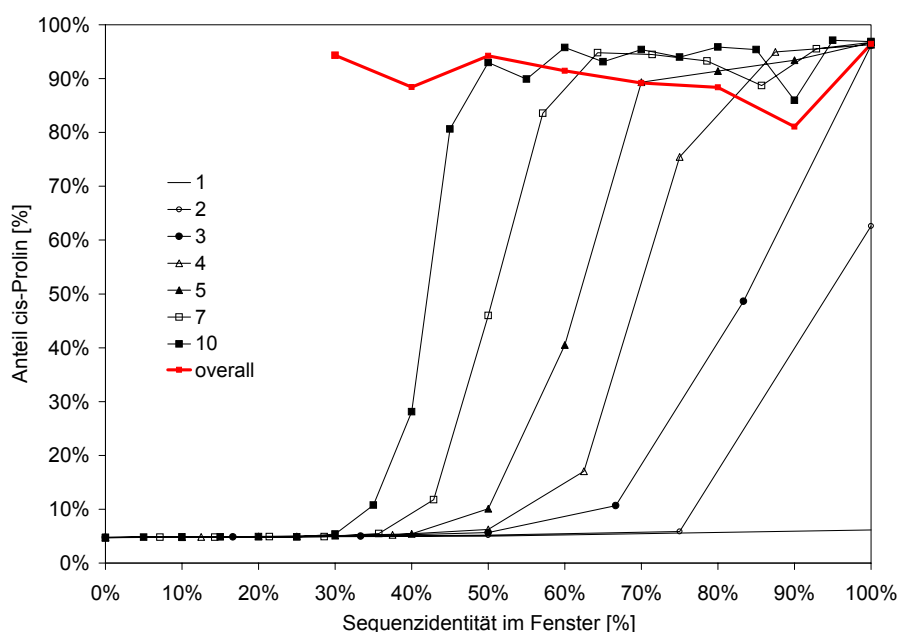


Abbildung 4.11: Anteil der Prolylreste in *cis*-Konformation für Peptide mit unterschiedlicher lokaler und globaler Sequenzidentität zu *cis*-Prolylresten bei verschiedenen Fenstergrößen der lokalen Sequenzidentität

Die Abbildung zeigt das Ausmaß der Konservierung der *cis*-Konformation in alignierten Proteinen in Abhängigkeit von der Sequenzidentität der Proteine.

Die Proteinfamilien wurden nach Vorkommen von *cis*-Prolinen und großer sequentieller Vielfalt ausgesucht, um auch strukturelle Alignments mit kleinen Sequenzidentitäten einzubeziehen.

Offensichtlich ist schon eine Sequenzidentität von ca. 20% ausreichend zur Konservierung der Konformation der Bindung. Unterhalb dieser Marke ist die *cis*-Konformation nicht mehr konserviert.

4.9 Sequenzalignments

Da die Vorhersage der Bindungskonformation aufgrund lokaler Sequenzumgebungen nicht zu befriedigenden Resultaten führte, wurden nun mit den Sequenzen aus dem Datensatz jeweils Sequenzalignments zu sämtlichen anderen in der PDB vorhandenen Proteinen durchgeführt, die die oben erwähnten Qualitätskriterien hinsichtlich Auflösung und R-Faktor erfüllen. Alignments eines Proteins mit sich selbst wurden nicht in die Bewertung einbezogen, um so die Situation zu simulieren, die Konformation einer Prolylbindung in

Tabelle 4.3: Superfamilien zur Untersuchung der strukturellen Alignments in Abb. 4.12.

Name	SCOP code	Proteine	Proline	<i>cis</i> -Proline
Restriktions-Endonucleasen	52980	144	1092	47
mikrobielle RNAsen	53933	242	1007	258
nucleotidbindend	51971	77	1009	52
Alkalische Phosphatasen	53649	77	1723	49

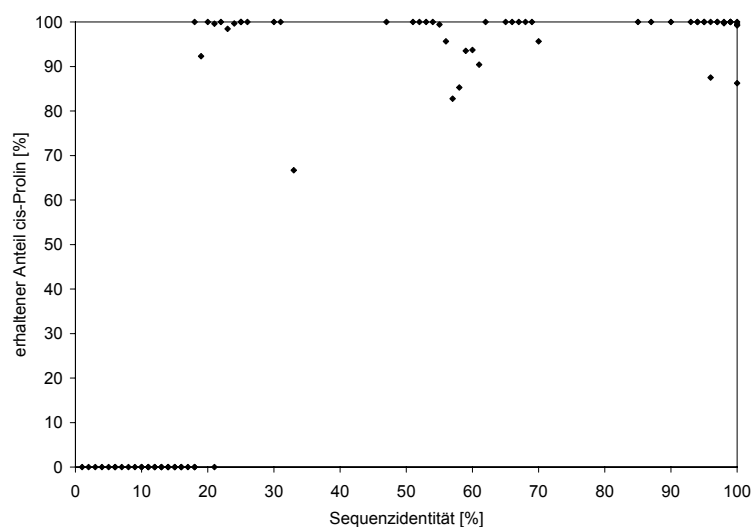


Abbildung 4.12: Konservierung von *cis*-Prolylresten in strukturellen Alignments. Proteine aus vier SCOP-Superfamilien (Tab. 4.3) wurden strukturell aligniert. Gezeigt ist der Anteil an *cis*-Prolylresten, der im Alignment mit *cis*-Prolylresten aligniert wird, in Abhängigkeit von der Gesamt-Sequenzidentität der Proteine.

einem bislang unbekannten Protein vorherzusagen.

Abbildung 4.13 zeigt den Anteil der korrekt identifizierten *cis*- und *trans*-Prolylreste, wenn jeweils das beste in der PDB gefundene Alignment als Vorhersagekriterium für die Konformation des Prolylrestes betrachtet wird. Unter Einbeziehung aller Strukturen außer dem Originalprotein können 62% der *cis*-Proline als solche erkannt werden, 98% der *Trans*-Proline werden richtig vorausgesagt.

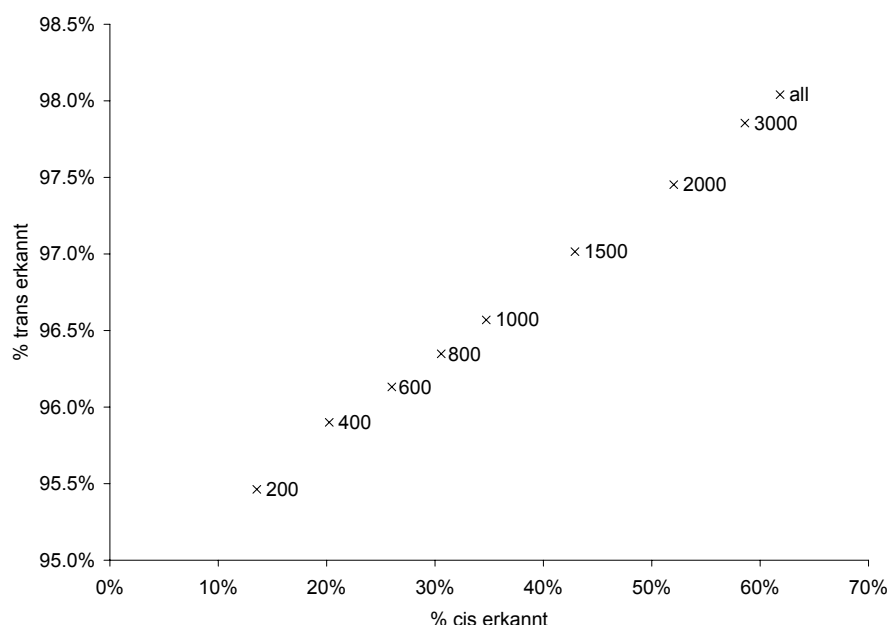


Abbildung 4.13: Vorhersagen der Konformation der Prolylbindung aufgrund von Sequenzalignments. Als Voraussage wurde jeweils die Konformation des Prolylrestes in der mit bestem score alignten Sequenz gewählt. Gezeigt sind die Erkennungsraten von *cis*- und *trans*-Prolylresten bei verschiedenen Obergrenzen der Güte des Alignments.

Um die Vorhersage unbekannter Proteine mit geringer Ähnlichkeit zu bekannten Proteinen zu simulieren, wurde die Vorhersage auch auf Grundlage von Alignments unterhalb der in Abb. 4.13 angegebenen Scores durchgeführt. Es fällt auf, daß die Vorhersagequalität in Bereichen, in denen die Familienzugehörigkeit noch problemlos ermittelt werden kann (*score* von ca. 200; vgl. Abb. 5.7), die Voraussage der Konformation der Prolylbindung schwieriger wird.

Statt nur das beste Alignment zur Auswertung heranzuziehen, kann auch versucht werden, aus mehreren guten Alignments eine konsistente Aussage zu treffen. Dazu wird ein gewichtetes Abstimmungsverfahren eingesetzt. Das

Gewicht jedes einzelnen Alignments sollte dabei vom Score abhängig sein, Alignments mit *cis*-Prolinen positiv und solche mit *trans*-Prolinen negativ zählen.

Zur Wichtung der Alignments wurde die Exponentialfunktion mit dem Score im Exponenten gewählt ($w = e^{\lambda \cdot \text{score}}$, Abb. 4.14). Bei der Wahl von $\lambda = 0,035$ nimmt der AUC-Wert ein Maximum von 0,915 an. Durch zusätzliche Information auch schlechterer Alignments kann also im Vergleich zur Modellierung nur nach dem besten Alignment noch eine Verbesserung erzielt werden. Große Werte für λ entsprechen einer stärkeren Wichtung der Alignments mit großen Score und sind daher eher äquivalent zur Modellierung auf Grundlage des besten Alignments.

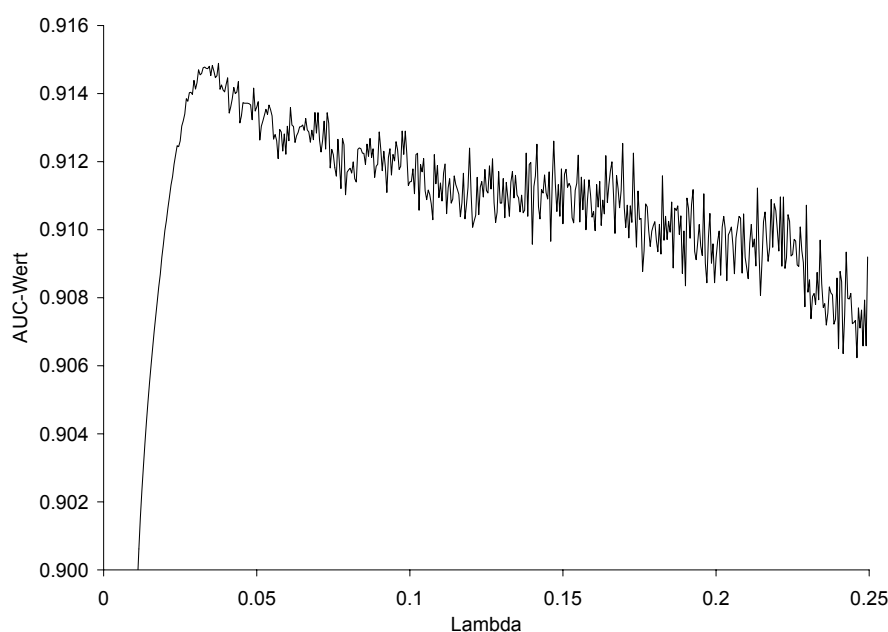


Abbildung 4.14: AUC-Werte der Vorhersage der Prolylbindung durch ein gewichtetes Abstimmungsverfahren. Gezeigt ist die Abhängigkeit des AUC-Wertes vom Parameter λ (s. Text).

4.10 Diskussion

Es konnte gezeigt werden, daß die Charakteristika der benachbarten Aminosäuren sich zwischen *cis*- und *trans*-Prolylresten unterscheiden. Wie bereits von anderen Autoren berichtet [39, 184, 175], tritt dieser Unterschied am deutlichsten bei der dem Prolylrest vorangehenden Aminosäure hervor.

An dieser Stelle sind Tryptophan und Tyrosin vor *cis*-Prolylbindungen stark überrepräsentiert.

In Übereinstimmung mit Pal und Chakrabarti [176] wurde gefunden, daß aromatische und kleine Aminosäuren sowie Prolin selbst an Position -1 *cis*-Prolylbindungen begünstigen und verzweigtkettige aliphatische Reste die Ausbildung von *cis*-Prolylbindungen erschweren.

Im Gegensatz zu Pal und Chakrabarti wurde aber kein positiver Einfluß polarer Aminosäuren festgestellt. Die Begünstigung von *cis*-Prolylbindungen nach aromatischen Aminosäuren könnte auf ein *stacking* der Ringe oder $\text{CH} \cdots \pi$ -Interaktionen zurückzuführen sein [176, 185].

Es konnte gezeigt werden, daß der Einfluß der Aminosäuren vor Prolylresten in Proteinen und Peptiden qualitativ identisch ist [177] – ein Hinweis darauf, daß die lokale Sequenzumgebung die *cis-trans*-Isomerisierung beeinflußt. Auch der kleine Glycylrest an dieser Stelle begünstigt die Ausbildung einer *cis*-Prolylbindung, was auf dessen sterische Flexibilität zurückzuführen ist. Auch Phenylalanin und Tyrosin, nicht aber Tryptophan begünstigen als Vorgänger des Prolylrestes eine *cis*-Peptidylbindung.

Eine Möglichkeit für das gehäufte Auftreten bestimmter Aminosäurereste vor *cis*-Prolinen könnte eine höhere Affinität der betreffenden Proteine zu Prolyl-Isomerasen (PPI) sein. Für die PPI Pin1 ist als eindeutiges Sequenzmotiv ein Phosphatidylserin/-Threonin vor dem Prolylrest entscheidend [163]. Für Parvulin wurden R, L, F [186] oder G, A, V, I, L, F, W, H, K und G [187] als bevorzugte Aminosäure vor Prolin berichtet, für FKBP's ein acht-Aminosäuren-Motiv mit vielen basischen und aromatischen Resten und zwei positiven Ladungen [188], und vor dem Prolin die Aminosäuren L, F, I [189]. Cyclosporine bevorzugen A, V, L, E, F oder G an der Position vor dem Prolylrest in Substraten [189]. Diese Sequenzmotive decken sich nicht mit den Aminosäuremotiven, die die Ausbildung von *cis*-Prolylresten fördern, eine kinetische Limitierung für die Ausbildung der *cis*-Prolylbindung kann daher nicht angenommen werden.

Während die Betrachtung der lokalen Sequenzumgebung allein keine verlässliche Vorhersage der Konformation der Prolylbindung ermöglicht, ist die Analyse breiterer Sequenzfenster informativer. Die verlässlichste Methode zur Vorhersage der Bindungskonformation bleibt das Sequenzalignment und daher die Untersuchung von Homologie zu anderen Proteinen.

Da auch die Betrachtung von Resten außerhalb der direkten Sequenzumgebung des Prolylrestes die Vorhersagequalität noch steigert, kann geschlossen werden, daß die Homologiebeziehungen zu anderen Proteinen und damit *long-range* Effekte und nicht nur die lokale Sequenzumgebung die Konformation der Prolylbindung festlegen.

Diese Feststellung wird durch den Befund bestätigt, daß Sequenzidenti-

täten von nur 20% schon ein verlässlicher Indikator für die Konformation der Prolylbindung darstellen. Im Homologie-Modelling ist daher ein Modellbau auf Grundlage des Proteins mit der größten Sequenzidentität ratsam [190].

Die Wichtigkeit von Prolylresten für Proteine wird auch in der Konservierungsanalyse deutlich: während Prolylreste an sich schon höher konserviert sind als der Rest der Proteine, übertreffen *cis*-Prolylreste die *trans*-Prolylreste noch im Grad ihrer Konservierung und wurden (nach Cystylresten) als der am zweithöchsten konservierte Rest beschrieben [190].

Zusätzlich zum Konservierungsgrad unterscheidet sich auch das Austauschprofil von *cis*- und *trans*-Prolylresten. Es konnte erwartet werden, daß der flexible Glycylrest ein bevorzugter Substituent für *cis*-Prolylreste darstellt, aber auch Cystein substituiert *cis*-Prolylreste häufiger als *trans*-Prolylreste. Offensichtlich sind einige Aminosäuren besser als andere in der Lage, die strukturelle Rolle von *cis*-Prolylresten zu übernehmen.

Der größte Unterschied tritt beim Alanin auf, das *trans*-Prolylreste in 16% der Fälle vertritt, während nur 10% der *cis*-Proline zu Alanin mutieren. Die häufige Mutation von *trans*-Prolin zu Alanin kann durch die Ähnlichkeit der Codons, die sich nur durch ein Nucleotid unterscheiden, erklärt werden, während die Bevorzugung des Glycins als Substituent für *cis*-Prolin auf dessen strukturelle Flexibilität zurückzuführen ist.

Die Untersuchung hat gezeigt, daß *cis*-Prolylreste wichtige Aufgaben in Proteinstrukturen erfüllen, einen größeren Grad an evolutionärer Stabilität und ein gegenüber *trans*-Prolylresten verändertes Mutationsprofil zeigen. Im Gegensatz zu lokalen Sequenzmotiven ist die Homologie zwischen Proteinen ein wesentlich besserer Indikator für die Konformation von Prolylresten.

Kapitel 5

Globale Sequenzähnlichkeiten

5.1 Sequenzähnlichkeit und Homologie

Eines der größten Probleme in der Biologie ist das Rätsel der Proteinfaltung. Trotz jahrzehntelanger intensiver Forschung ist allerdings der genaue Mechanismus der Faltung noch unbekannt.

Um zu einer gegebenen Sequenz ein strukturelles Modell des Proteins zu erhalten, gibt es die Möglichkeit des *homology modelling* und *threading* (zur Übersicht s. [191]), die beide auf Vergleichen mit bereits bekannten Proteinstrukturen beruhen. Beim *homology modelling* wird ein Modell des Proteins nach dem Muster eines Proteins bekannter Struktur und ähnlicher Sequenz gebildet, beim *threading* wird die zu untersuchende Sequenz im Computer auf die Struktur vieler bekannter Proteine abgebildet und die Wahrscheinlichkeit der Richtigkeit des Modells anhand von berechneten Faltungsenergien beurteilt. Von Proteinen mit hinreichender Sequenzähnlichkeit zu einem bekannten Protein kann angenommen werden, daß sie sich ähnlich wie das Referenzprotein falten [29].

Dies beruht darauf, daß man bei hinreichender Sequenzähnlichkeit von Proteinen davon ausgehen kann, daß diese evolutionär verwandt sind. Im Verlauf der Evolution bleibt die Faltung des Proteins erhalten, um dessen Funktion zu gewährleisten. Dies führt dazu, daß an den für die Faltung essentiellen Positionen nur Aminosäuren toleriert werden, die eine ähnliche Faltung der Peptidkette hervorrufen.

Über den Einfluß von Sequenzähnlichkeiten ohne evolutionären Hintergrund auf Ähnlichkeiten in der Faltung von Proteinen ist wenig bekannt. Man kennt allerdings einige Beispiele für stark ähnliche Sequenzen ohne entsprechende Ähnlichkeiten in der Faltung (sog. *Chamäleon-Sequenzen*, s. S. 18). Ein anderes Beispiel für Sequenzähnlichkeiten ohne evolutionären Hin-

tergrund ist die inverse Sequenzähnlichkeit zwischen Proteinen, die im Folgenden untersucht werden soll.

5.2 Hat inverse Sequenzähnlichkeit eine Bedeutung?

Für eine strukturelle Bedeutung von inverser Sequenzähnlichkeit könnte es mehrere Gründe geben – bei einer Invertierung der Sequenz bleiben die Aminosäurezusammensetzung und Nachbarschaften zwischen Aminosäuren bestehen. Das Hydropathieprofil bleibt ebenfalls erhalten. Auch Periodizitäten, die zum Beispiel bei der Ausbildung von amphipathischen Helices eine Rolle spielen, sind unverändert. Aufgrund dieser Eigenschaften wurde spekuliert, daß invertierte Sequenzen sich zu einem Spiegelbild des originalen Proteins falten, da das „Vertauschen“ von N- und C-terminalen Nachbarn einer Aminosäure einer Spiegelung ähnelt [192].

Strukturvergleiche von kurzen Proteinabschnitten, die in beiden Durchlaufrihtungen in der PDB vorkommen, zeigten häufig eine strukturelle Ähnlichkeit von Vorwärts- und inversen Sequenzen [30]. Aufgrund computergestützter *lattice model* Studien an retro-Protein A kamen Olszewski *et al.* [193] zu der Überzeugung, das invertierte Protein müsse die gleiche Faltung wie das Originalprotein besitzen. Eine gespiegelte Struktur wird von den Autoren auch aufgrund des Auftretens lingsgängiger Helices für unwahrscheinlich gehalten.

Die B-Domäne von Protein A, eine SH3-Domäne und die B1-Domäne von Protein G erwiesen sich invertiert jedoch als ungefalted [194], ebenso wie synthetisches (C. Frömmel, persönl. Mitteilung) oder in *E. coli* exprimiertes (H. Böcker, persönl. Mitteilung) invertiertes Ubiquitin. Nach Zugabe von Trifluoroethanol, das α -Helices stabilisiert, ist die invertierte B-Domäne von Protein A jedoch stabil (ohne Zugabe von Trifluorethanol war selbst die ursprüngliche Domäne ungefalted). CD-Spektroskopie zeigte, daß das invertierte Protein eine dem Vorwärtsprotein ähnliche Sekundärstruktur-Zusammensetzung und damit wahrscheinlich eine ähnliche Faltung aufweist [195].

Der *leucine zipper* GCN4 wurde ebenfalls invers synthetisiert und kristallisiert [196, 197]. Das invertierte Protein bildet stabile Tetramere, und seine Struktur ähnelt sehr stark der des Vorwärtsproteins – die C_α -Atome der beiden Strukturen lassen sich mit einem RMSD von 0.37 Å überlagern.

Als Beispiel für ein teilinvertiertes Protein kann α -Hämolyisin mit invertierter, 25 Aminosäuren langer Transmembrandomäne angeführt werden. Das Protein ist funktionell und bildet Poren mit ähnlichen Eigenschaften wie das

Originalprotein [198].

Die Symmetrie in der Sequenzinformation wird auch durch den Befund deutlich, daß Sekundärstruktur-Vorhersageprogramme in den meisten Fällen für invertierte Sequenzen die gleiche Vorhersage liefern wie für Vorwärtssequenzen. Stimmt die Vorhersage überein, steigt die Vorhersagequalität um 4% [199].

Im Gegensatz zu Proteinen ist über kleinere Peptide mit invertierten Sequenzen wesentlich mehr bekannt. *Retro-inverso*-Peptide (invertierte Peptide mit D- statt L-Aminosäuren) sind dem Originalpeptid strukturell meist sehr ähnlich und werden seit langem als Peptidomimetika verwendet (zur Übersicht s. [200, 201]). Sie besitzen gegenüber dem Ausgangspeptid den Vorteil einer erheblich größeren Stabilität, zum Beispiel gegen Proteasen. Beispiele von Peptiden, deren *retro-inverso*-Form als Peptidomimetikum benutzt wird, sind Dermorphin, Enkephalin, Neurotensin und Somatostatin. Besonders kleine porenbildende Peptide sind auch in ihren *inverso*-, *retro*- und *retro-inverso*-Form funktionell. Beim Melittin hat die *Retro*-Form auch den Vorteil, daß die unerwünschte hämolytische Wirkung bei gleichbleibender antimikrobieller Wirkung entfällt [202]. *Retro-Inverso*-Analoga von Peptiden binden auch Antikörper gegen das Originalpeptid und lassen sich im Gegenzug auch zur Herstellung von Antikörpern gegen das Ausgangspeptid verwenden [203].

Viele Autoren fanden, daß Peptide und auch ganze Proteine aus D-Aminosäuren die spiegelbildliche Struktur zu den entsprechenden L-Proteinen ausbilden (zur Übersicht s. [204]). Ein schlagender Beweis wurde durch die Totalsynthese von D-HIV-Protease geführt – das Protein zeigt reziproke chirale Spezifität zu Substraten und Inhibitoren wie das Originalprotein [205].

Die generelle Frage, ob invertierte Proteine eine definierte Struktur besitzen, und ob diese Struktur der des Vorwärtsproteins ähnelt, bleibt jedoch bestehen.

5.3 Zielsetzung

Anhand von Sequenzvergleichen und strukturellen Überlagerungen von Proteinen bekannter Struktur soll herausgearbeitet werden, inwieweit eine inverse Sequenzähnlichkeit auch eine Ähnlichkeit der Struktur impliziert. Auch die Hypothese, daß invertierte Proteine das Spiegelbild des Originalproteins bilden, soll überprüft werden.

5.4 Die PDB als Datenquelle

Die *Protein Data Bank* (PDB) enthielt zum Untersuchungszeitpunkt ca. 20 000 Proteinstrukturen. Dabei ist allerdings zu bedenken, daß besonders interessante oder leicht zu kristallisierende Proteine in mehreren Strukturen vorliegen und dadurch eine nichtrepräsentative Häufung dieser Proteine (*bias*) erzeugt wird. Für die Untersuchung werden daher nichtredundante Protein-Datensätze benötigt, die durch den PISCES-Server [179] bereitgestellt werden. Die Datensätze bestehen jeweils aus Proteinen mit einer Auflösung von $\leq 2,5$ Å und einem R-Faktor von $\leq 0,25$, deren Sequenzidentität untereinander einen gewählten Schwellenwert unterschreitet. Je restriktiver (kleiner) der Schwellenwert, desto kleiner ist der erhaltene Datensatz.

5.5 Inverse Sequenzalignments

Sequenzalignments sämtlicher Protein-Datensätze gegen sich selbst in gleichlaufender und entgegengesetzter Durchlaufrichtung wurden mittels BLAST [28] mit Standardparametern (BLOSUM62 Matrix [24, 206], *gap penalty* 11, *gap extension penalty* 1) durchgeführt. Treffer mit einem E-Wert unter 10 wurden als potentiell signifikant gewertet. Alignments eines Proteins mit sich selbst wurden nicht mitgezählt.

Zunächst fällt auf, daß bei restriktiverer Wahl des Sequenzidentitätsschwellenwertes im Protein-Datensatz die Anzahl der Vorwärtsalignments im Vergleich zu den erhaltenen inversen Alignments abnimmt (Abb. 5.1).

Während eine Vergrößerung des Protein-Datensatzes durch Zulassen größerer Sequenzidentitäten innerhalb der Proteine im Datensatz zu einem linearen Anstieg der inversen Alignments führt, steigt die Zahl an Vorwärtsalignments wesentlich schneller an. Dies ist ein Indiz für das Vorkommen von überrepräsentierten Gruppen homologer Proteine, die untereinander eine Vielzahl von Alignments ermöglichen (*bias*).

Ab einem Schwellenwert von 70% steigt die Anzahl dieser Gruppen rapide an. Um also mit einem möglichst nichtredundanten Datensatz zu arbeiten, der dennoch hinreichend viele Proteine enthält, erscheint ein Schwellenwert von 50% gegenseitiger Sequenzidentität sinnvoll – die Zunahme an Vorwärtsalignments ist in diesem Bereich noch annähernd linear, was für eine nur geringe Häufung von Gruppen homologer Proteine spricht. Der Datensatz enthält bei der Wahl dieses Schwellenwertes 3904 Proteine.

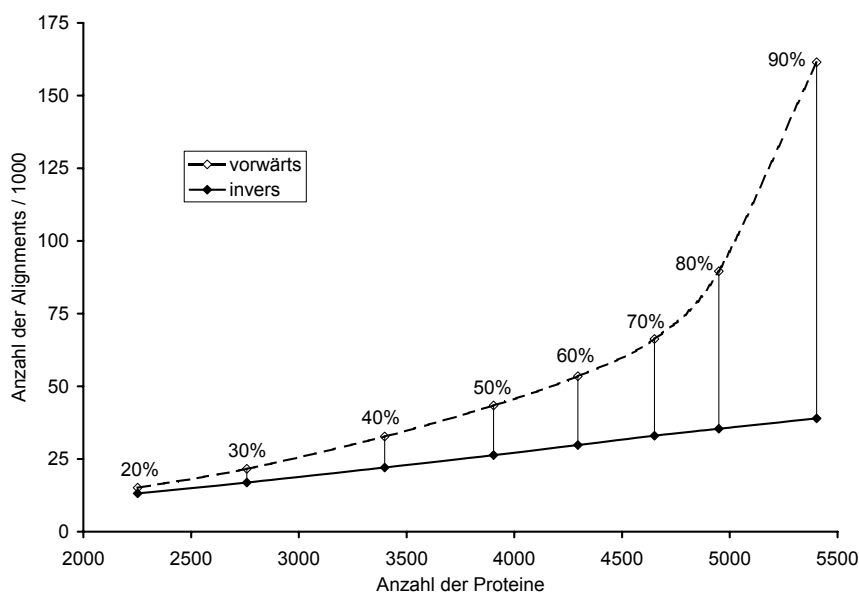


Abbildung 5.1: Anzahl der Vorwärts- und Rückwärtsalignments bei verschiedenen Schwellenwerten der Sequenzidentität im Protein-Datensatz. Die Zahl der Vorwärtsalignments sinkt bei restriktiverem Schwellenwert schneller als die Zahl der inversen Alignments. Über den Datenpunkten ist der Schwellenwert der Sequenzidentität im jeweiligen Datensatz angegeben.

5.6 Alignments mit Zufallssequenzen

Wenn in Proteinsequenzen lokale Häufungen ähnlicher Aminosäuren auftreten, sollte die gefundene Anzahl an Alignments mit invertierten Proteinen größer als mit Zufallssequenzen sein. Daß derartige Häufungen auftreten, ist seit langem bekannt: so finden sich zum Beispiel in Transmembranhelices vermehrt hydrophobe Reste, in α -Helices und β -Sheets herrschen ebenfalls bestimmte Reste vor, die zur Ausbildung der jeweiligen Sekundärstruktur beitragen. Im Inneren von Proteinen häufen sich ebenfalls hydrophobe Reste, während die Außenfläche eher von polaren Aminosäuren gebildet wird.

Diese Häufungen lassen sich mittels Markov-Modellen von Sequenzen simulieren. Das einfachste Beispiel eines Markov-Modells ist ein Sequenzgenerator, der unabhängig von der vorangehenden Aminosäure stets entsprechend der Häufigkeit der Aminosäuren im Proteom eine weitere Aminosäure an die wachsende Kette anhängt. Bei Erhalt der Häufigkeiten von Aminosäurepaaren ergibt sich ein Sequenzgenerator, der die Wahl der nächsten angefügten Aminosäure jeweils von der vorangehenden Aminosäure abhängig macht. Auf

die selbe Weise lassen sich Sequenzen erzeugen, deren Tripel- oder Quadrupelhäufigkeit usw. realen Sequenzen entspricht.

Um der Frage nachzugehen, inwieweit derartige Häufungen für das Zustandekommen inverser Alignments verantwortlich sind, wurden neue Protein-Datensätze durch Durchmischen der Sequenzen und Markov-Generatoren erzeugt. Anzahl und Länge der Proteine in diesen Datensätzen entsprach dabei dem Originaldatensatz. Die erhaltenen Protein-Datensätze wurden gegen den ursprünglichen Datensatz aligniert (Tab. 5.1).

Tabelle 5.1: Anzahl von Alignments gegen eine Teilmenge der PDB mit 50% Sequenzidentitäts-Schwellenwert. *shuffle*: jede einzelne Sequenz wurde durchmischt. *MM*: durch Markov-Modelle generiert.

Datensatz	Alignments
shuffle	21 761
MM	20 791
MM (Paare erhalten)	21 669
MM (Tripel erhalten)	23 130
MM (Quadrupel erhalten)	24 469
invers	27 295
vorwärts	47 303

Die Anzahl inverser Alignments beträgt 125% der Anzahl von Alignments mit durchgemischten Sequenzen. Auch unter Beibehaltung der Häufigkeiten von Tupeln ist die Zahl gefundener Alignments geringer als bei Invertierung der Sequenz. Nach Invertierung sind die erhaltenen Sequenzen also „proteinähnlicher“ als bei Erhalt der Tupel.

5.7 Generierung der Alignment-Datensätze

Um unter den inversen Alignments keine Häufung „einfacher“ Sequenzen (wie z. B. AGAGAGA) zu erhalten, die sich auch in Vorwärtsrichtung ähneln, wurden sämtliche invers alignierte Proteinpaare nochmals vorwärts aligniert. Paare, bei denen der Score eines Vorwärtsalignments den des inversen Alignments übertraf, wurden aus dem Datensatz entfernt. Der endgültige Alignment-Datensatz enthielt 26 315 inverse Alignments.

Um Vorwärts- und Rückwärtsalignments besser vergleichen zu können, sollte die Verteilungen der Länge und des Alignment-Scores bei beiden möglichst gleich sein. Daher wurde aus der Gruppe der Vorwärtsalignments eine

Teilmenge gebildet, indem für jedes inverse Alignment das Vorwärtsalignment mit der geringsten Differenz in Score und Länge in den Datensatz übernommen wurde. Dieser enthält somit jeweils die gleiche Anzahl von Alignments und die gleiche Score- und Längenverteilung wie der Datensatz der Rückwärtsalignments (Abb. 5.2).

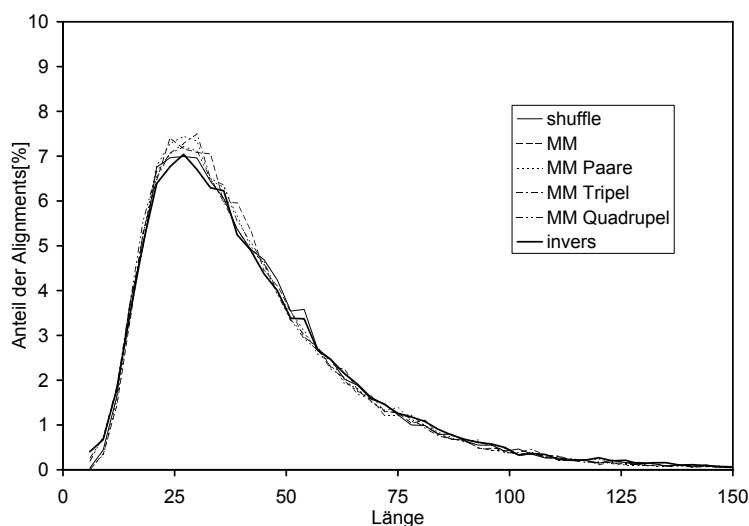


Abbildung 5.2: Längenverteilung bei Alignments verschiedener Sequenzen mit PDB-Daten. Die Verteilungen für inverse, geschuffelte und Markov-Modell-Alignments sind identisch.

5.8 Strukturelle Überlagerungen

Um die Hypothese zu überprüfen, daß inverse Sequenzalignments eine strukturelle Ähnlichkeit implizieren, wurden die alignierten Proteine überlagert [19] (Abb. 5.3 und 5.4). Dabei wird die Translations- und Rotationsbewegung ermittelt, nach deren Anwendung die Summe der Abstandskvadratrate (RMSD) der miteinander alignierten $C\alpha$ -Atome minimal wird. Um die Hypothese zu prüfen, daß ein invertiertes Protein gegebenenfalls das Spiegelbild der Originalstruktur bildet [192], wurden die invers alignierten Proteine auch mit dem Spiegelbild der Originalproteine überlagert (Abb. 5.5).

Obwohl die Verteilung der Längen und Scores bei Vorwärts- und Rückwärtsalignments gleich ist, finden sich bei den Vorwärtsalignments in Abb. 5.3 zwei deutlich voneinander getrennte Punktwolken, die durch eine Trennlinie ($rmsd=0.1 \cdot len$) separiert werden. 1731 von 19248 Alignments fallen bei den

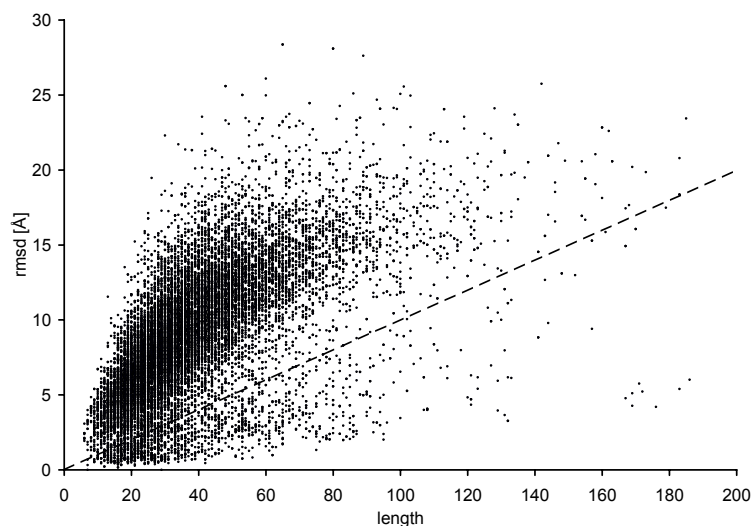


Abbildung 5.3: RMSDs der überlagerten vorwärts alignierten Proteine. Eine Aufteilung in strukturell bedeutungsvolle Alignments (kleine RMSDs) und offensichtlich „falsche“ Alignments (große RMSDs) ist deutlich zu erkennen.

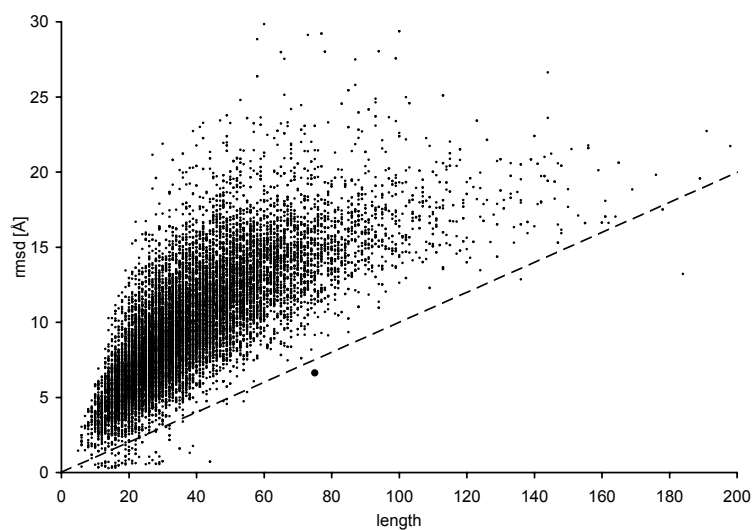


Abbildung 5.4: RMSDs der überlagerten invers alignierten Proteine. Hier fehlt die Gruppe der strukturell bedeutsamen Alignments mit Längen über 50 Aminosäuren. Der hervorgehobene Datenpunkt markiert eine im Text behandelte Ausnahme.

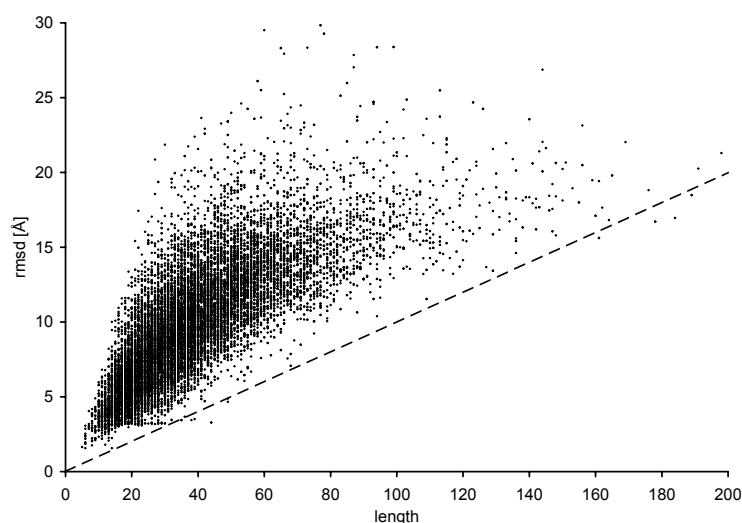


Abbildung 5.5: RMSDs der invers alignierten Proteine mit Spiegelbildern der Originalproteine. Auch hier fehlt die Gruppe der strukturell bedeutsamen Alignments.

Vorwärtsalignments unter die Trennlinie, bei den inversen (Abb. 5.4) nur 99 von 19 073. Bei den gespiegelten Überlagerungen gibt es nur 7 Datenpunkte unterhalb der Trennlinie. Offensichtlich sind die Vorwärtsalignments also zumindest teilweise ein Indiz für strukturelle Ähnlichkeit, inverse Alignments hingegen trotz gleicher Qualitätsmerkmale des Alignments nicht.

5.9 Familienzugehörigkeiten alignierter Proteine

Sind die kleinen RMSD-Werte bei der Überlagerung der vorwärts alignierten Proteine ein Indiz für Familienzugehörigkeiten? Es wäre durchaus denkbar, daß durch das Alignment nur kleine strukturell ähnliche Bereiche gefunden werden (zum Beispiel längere Helices, HTH-Domänen), ohne daß die beteiligten Proteine miteinander verwandt sind. Um dieser Frage nachzugehen, wurden Proteine aus der SCOP-Datenbank [207, 183] mit höchstens 40% paarweiser Sequenzidentität von der ASTRAL-Datenbank [208] bezogen und untersucht.

Die SCOP-Datenbank bietet den Vorteil, daß sämtliche enthaltene Proteine in ein System von Klassen, Superfamilien und Familien eingeteilt sind. Dadurch ist es möglich, den Grad an Verwandtschaft zwischen zwei Proteinen durch einen Vergleich ihrer Zugehörigkeiten zu den einzelnen Klassen zu

ermitteln. Der Datensatz aus SCOP enthält 3600 Proteinketten. Nach Alignierung durch BLAST wurden die einander zugeordneten Sequenzabschnitte überlagert und deren RMSD bestimmt (Abb. 5.6).

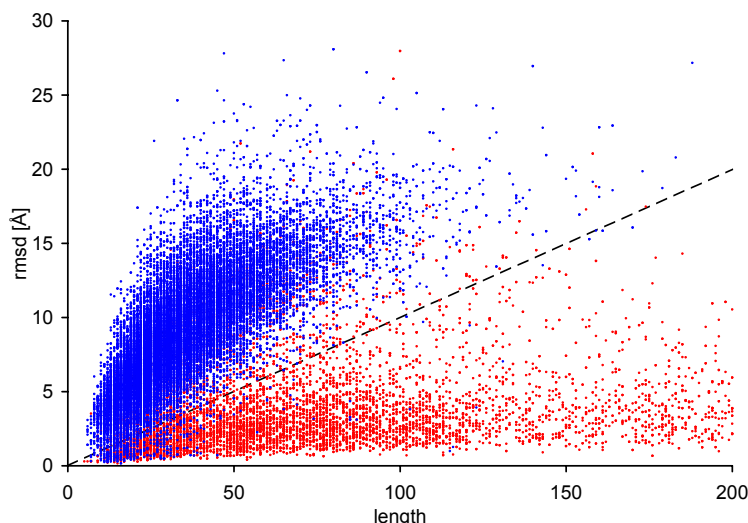


Abbildung 5.6: RMSDs der alignierten Proteine aus der SCOP-Datenbank. Rot: identische Superfamilie; blau: unterschiedliche Superfamilie

Die Datenmenge bildet, ebenso wie die Vorwärtsalignments der PDB-Proteine, zwei deutlich voneinander getrennte Teilmengen. Die schon bei der Untersuchung der PDB-Proteine gefundene Trennlinie separiert auch hier die beiden Teilmengen. Proteine, die mit kleinem RMSD überlagert werden können, gehören meist zur gleichen Superfamilie. Die Trennlinie identifiziert 6 996 von 8 369 (83,6%) der Proteine der gleichen Superfamilie und 28 109 von 28 549 (98,5%) der Proteine unterschiedlicher Superfamilien. Die in den Abbildungen benutzte Trennlinie ist somit geeignet, verwandte Proteine von unverwandten zu trennen.

Ist inverse Sequenzähnlichkeit ein Indikator für die Familienzugehörigkeit von Proteinen? Um diese Frage weiter zu klären, wurde eine nichtredundante Teilmenge der SCOP-Datenbank mit weniger als 50% paarweiser Sequenzidentität gleich- und gegenläufig aligniert. Wie aus Abb. 5.7 ersichtlich, ist für Vorwärtsalignments schon ein Score von 250 ein guter Indikator für die Zuordnung zu Proteinfamilien, während invers ähnliche Proteine nur in Ausnahmefällen der gleichen Familie angehören.

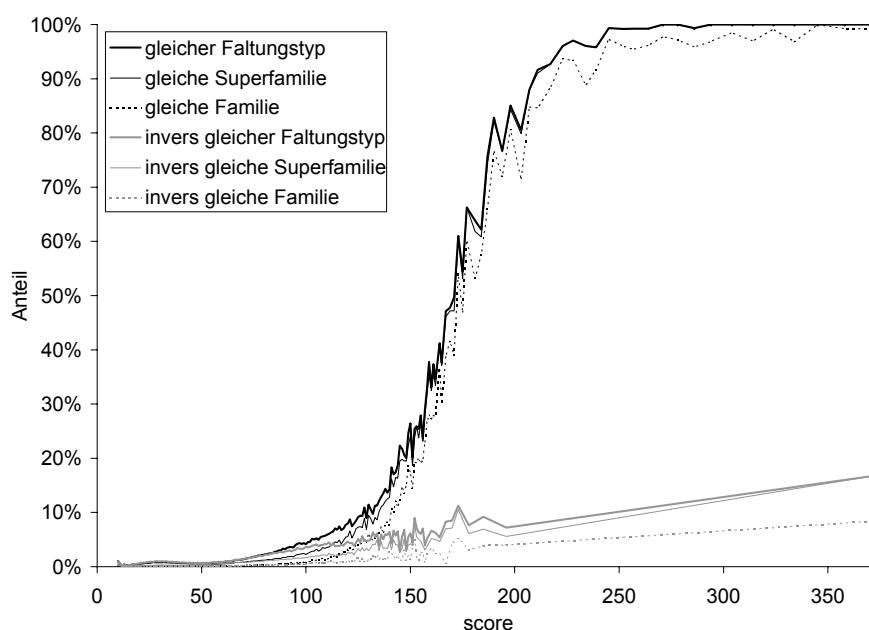


Abbildung 5.7: Anteil von vorwärts und invers alignierten Proteinen, die zur selben Faltungsklasse, Superfamilie oder Familie gehören, in Abhängigkeit vom Score des Alignments.

5.10 Strukturelle Verwandtschaft

Um zu untersuchen, ob Proteinpaare unterhalb der Trennlinie auch bei den PDB-Alignments verwandte Proteine umfassen, wurden derartige Paare unter den Vorwärts- und Rückwärtsalignments einer umfangreichen visuellen Prüfung unterzogen. Es zeigte sich, daß die betreffenden vorwärts alignierten Proteine zum größten Teil homolog und damit evolutionär verwandt sind oder zumindest gemeinsame Domänen umfassen, während die invers alignierten Proteine unterhalb der Trennlinie strukturell und auch funktionell unterschiedlich sind.

72 der 99 Paare invers alignierter Proteine unterhalb der Trennlinie stellen einfache helikale Strukturen mit einer Länge von bis zu 49 Aminosäuren dar, nur sieben Datenpunkte liegen im Größenbereich von Proteindomänen (ab 50 Aminosäuren). Zwei dieser Paare umfassen Proteine, die sich nur in ihrer Form ähneln: 1,4- β -D-Xylan-Xylanohydroxylase (1XYZ) und ein Transkriptionsfaktor (1HZ4) sowie Alkalische Phosphatase (1KAP) und Rhamnogalacturonase A (1RMG). Die beiden ersten Proteine bilden hufeisenförmige Strukturen, die zweiten längere Zylinder aus β -Sheets. In beiden Fällen stimmen die einander zugeordneten Sekundärstrukturen nicht überein.

Ein Beispiel eines strukturell ähnlichen Proteinpaars mit inverser Sequenzähnlichkeit (hervorgehoben in Abb. 5.4) ist in Abb. 5.8 gezeigt. Es handelt sich um Teile der Aldolase (PDB-Code 1EUA) mit Glycerol-Dehydrogenase (PDB-Code 1JQ5). Die beiden Proteine gehören unterschiedlichen Faltungstypen an und sind auch funktionell nicht verwandt.

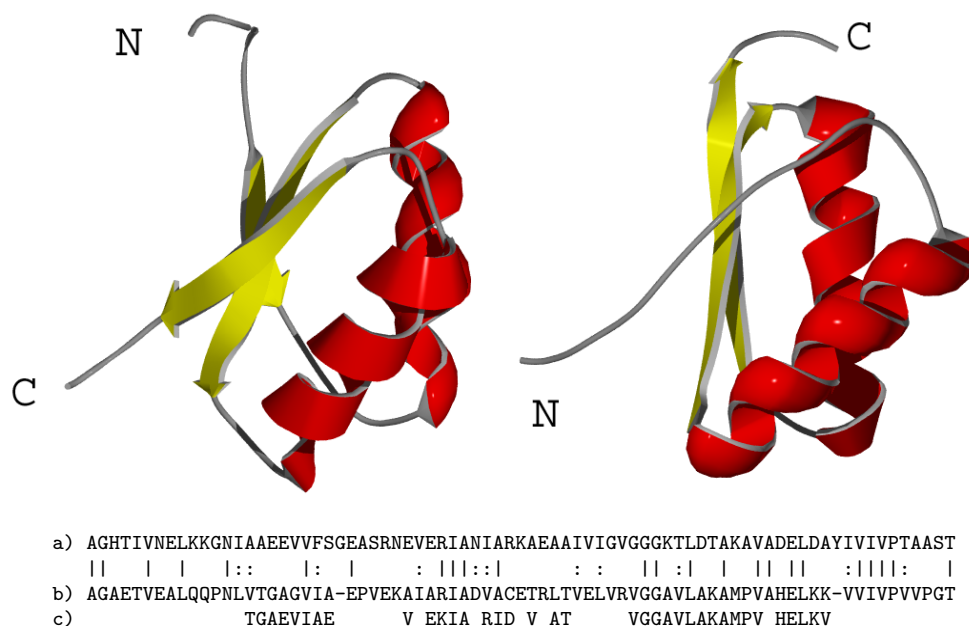


Abbildung 5.8: Teile der Aldolase (1EUA) zeigen inverse Strukturähnlichkeit mit einem Abschnitt der Glycerol-Dehydrogenase (1JQ5). Die Durchlaufrichtung des Peptid-Rückgrats ist dabei entgegengesetzt. a: Sequenz der Aldolase; b: invers alignierte Glycerol-Dehydrogenase; c: Überlagerte C α -Atome der Glycerol-Dehydrogenase

Dennoch findet der Algorithmus zur strukturellen Überlagerung ohne Sequenzinformation [209], daß die beiden Proteine in unterschiedlicher Durchlaufrichtung überlagert werden müssen (strukturelles Alignment in Abb. 5.8). 41 der 76 C α -Atome können mit einem RMSD von 1.73 Å überlagert werden.

Ein visueller Vergleich der anderen Datenpunkte unterhalb der Trennlinie in Abb. 5.4 zeigt jedoch, daß diese inverse strukturelle Verwandtschaft einen Ausnahmefall darstellt. Andere mit kleinem RMSD überlagerte Proteine zeigen trotz kleinem RMSD keine ersichtliche strukturelle Verwandtschaft und gehören auch nur in Ausnahmefällen gleichen Familien an.

5.11 Unterschiede zwischen Alignments

Die Charakteristika der Alignments zu beiden Seiten der Trennlinie unterscheiden sich nicht hinsichtlich ihres Scores im Alignment. Um Unterschiede zwischen den „bedeutungsvollen“ und „nicht bedeutsamen“ Alignments zu ermitteln, wurde das Auftreten von längeren Bereichen innerhalb des Alignments untersucht. Interessante Eigenschaften wären zum Beispiel konservierte oder ähnliche Bereiche oder Bereiche mit Aminosäurefolgen gleicher Eigenschaften (klein, geladen, hydrophob usw.). Beispielhaft ist in Abb. 5.9 die Häufigkeit von hydrophoben Bereichen steigender Länge in beiden Gruppen aufgezeigt. Als hydrophobe Bereiche zählen dabei Abschnitte, in denen in beiden alignierten Sequenzen hydrophobe Aminosäuren auftreten.

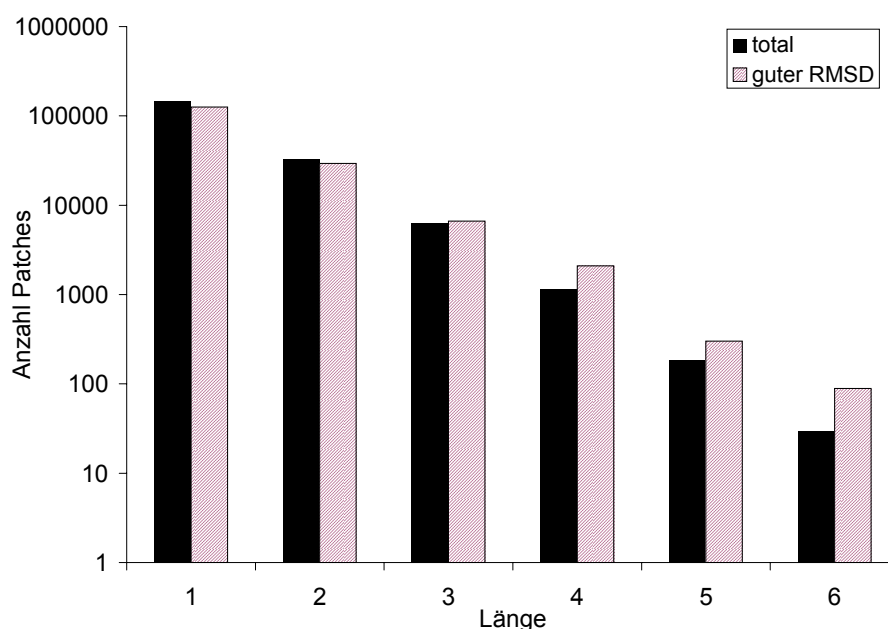


Abbildung 5.9: Längenverteilung hydrophober Patches in allen und „guten“ Alignments. Längere hydrophobe Bereiche sind in Alignments mit kleinem RMSD überrepräsentiert.

Wie man erkennt, sind zwar längere Bereiche hydrophober Aminosäuren in strukturell bedeutsamen Bereichen überrepräsentiert, der Effekt reicht jedoch nicht für eine Beurteilung einzelner Alignments aus. Ähnliche Befunde liefern auch die Untersuchungen anderer Bereiche. Auch inverse Alignments unterscheiden sich in diesen Kriterien nicht von Vorwärtsalignments.

5.11.1 Zuordnung von Sekundärstrukturen

Neben einer strukturellen Überlagerung ist auch die Zuordnung von Sekundärstrukturen von großem Interesse. Vorhersageprogramme liefern in den meisten Fällen identische Resultate für invertierte und „Vorwärtsproteine“ [199]. Stimmen auch invers alignierte Proteine im gleichen Maße wie vorwärts alignierte Proteine mit der Sekundärstruktur des Targetproteins überein?

Um diese Frage zu überprüfen, wurde zu den Proteinen aus der PDB mittels DSSP [210] die Sekundärstruktur-Zuordnung bestimmt. Die Sekundärstruktur vorwärts und invers alignierter Proteine wurde mit den Sekundärstrukturen des Originalproteins verglichen (Abb. 5.10).

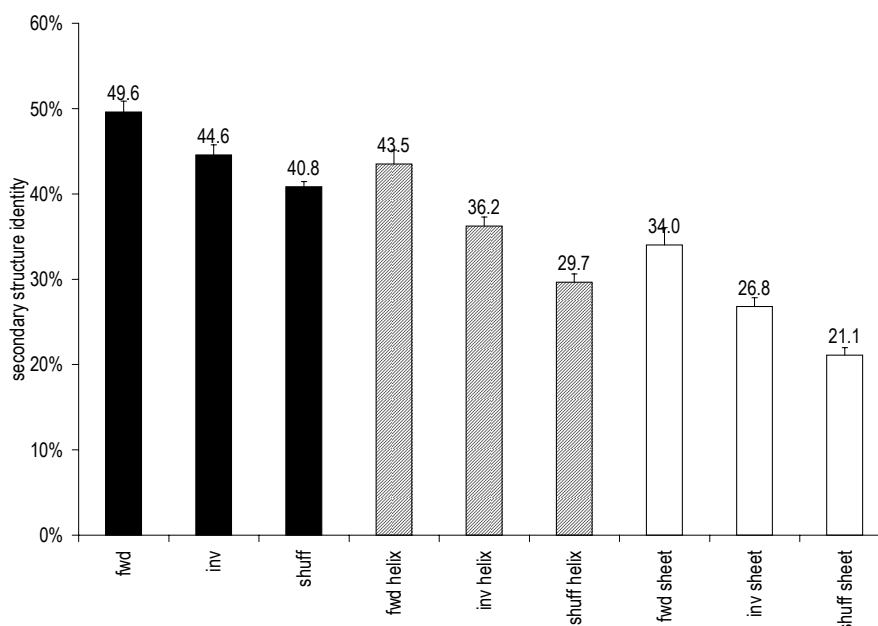


Abbildung 5.10: Übereinstimmung der Sekundärstrukturen vorwärts und invers alignierter Proteine mit dem Targetprotein. Inverse Alignments erlauben eine bessere Zuordnung als geschuffelte Alignments (Zufallszuordnung).

Die Abbildung zeigt den Prozentsatz korrekter Sekundärstrukturzuordnungen, wenn diese von Vorwärts-, inversen oder randomisierten Alignments übernommen werden. Eine randomisierte Zuordnung entsprach dabei einem zufälligen Mischen der mit den einzelnen Aminosäuren assoziierten Sekundärstrukturen.

Die Güte der Zuordnung mittels inverser Alignments liegt sowohl in der Gesamtwertung (schwarze Balken) als auch bei Betrachtung von α -Helices

oder β -Sheets zwischen der Zuordnung mittels Vorwärtsalignments und randomisierter Zuordnung.

Zu beachten ist dabei, daß die Verteilung der Scores und Längen bei Vorwärts- und Rückwärtsalignments identisch ist. Die Trefferquote der randomisierten Zuordnung beruht dabei nur auf Häufungen innerhalb von Proteinen (Proteine mit insgesamt hohem Anteil an α -Helices oder β -sheets).

Die Sekundärstrukturzuweisung aufgrund inverser Sequenzähnlichkeit zwischen Proteinen kann zum Teil auf die Häufung von sekundärstruktur-begünstigenden Aminosäuren (zum Beispiel Alanin, Glutamat oder Leucin in α -Helices) zurückzuführen sein. Auch die Beibehaltung von Periodizitäten, die zum Beispiel charakteristisch für amphiphatische Helices sind, trägt zur Ähnlichkeit in den Sekundärstrukturen von Proteinen mit inverser Sequenzähnlichkeit bei.

5.12 Diskussion

Obwohl inverse Sequenzähnlichkeit zwischen Proteinen kein seltenes Phänomen darstellt, konnte trotz der Vielzahl von Beispielen keine Beziehung zur strukturellen Ähnlichkeit hergestellt werden. Interessanterweise zeigen viele Proteine selbstinverse Ähnlichkeit, die oft auf die Wiederholung von kleineren Domänen oder Periodizitäten in Sekundärstrukturelementen zurückzuführen ist. Selbstinvers ähnliche Proteine besitzen oft eine stark symmetrische Struktur, zum Beispiel ring- oder faßförmig.

Proteine mit gegenseitiger inverser Sequenzähnlichkeit, deren Sequenzen sich in gleichlaufender Richtung nicht ähneln, zeigen meist keine Ähnlichkeit in den Strukturen. Als Ausnahme treten hier kurze helikale Bereiche auf, die teilweise durch eine gleichgerichtete Alignierung der Sequenzen nicht gefunden worden wären.

Die Untersuchung der Kriterien von Vorwärts- und Rückwärtsalignments zeigte keine Unterschiede zwischen den beiden Gruppen. Die Längenverteilung hydrophober Patches und die Verteilung der *gaps* unterscheidet sich nicht hinreichend, um zwischen strukturell bedeutsamen und unbedeutenden Alignments zu unterscheiden.

Obwohl die inversen Alignments zu einer schlechteren Strukturüberlagerung führten als die Vorwärtsalignments, konnten einige Helices mit inverser Sequenzidentität aufgefunden werden. Im Gegensatz zur globalen Faltung ist offenbar die räumliche Struktur der Sekundärstruktur-Elemente bei invertierten Sequenzen eher konserviert.

Der Grund für das gehäufte Auftreten inverser Sequenzähnlichkeiten ist unbekannt. Ein Grund wird sicher darin liegen, daß Cluster ähnlicher Ami-

nosäuren, die in Proteinen oft zusammen auftreten, erhalten bleiben, ebenso wie Periodizitäten, die bei der Ausbildung von α -Helices und β -Sheets eine Rolle spielen.

Das Fehlen von größeren invers ähnlichen Proteinpaaren mit ähnlichen Strukturen läßt vermuten, daß eine Invertierung der Sequenz zu ungefalteten Proteinen oder Proteinen mit gegenüber dem Originalprotein veränderter Faltung führt. Die Invertierung von Proteinsequenzen kann daher als eine Methode zur Erzeugung von *proteinähnlichen* Sequenzen unter Beibehaltung der Aminosäurezusammensetzung, Nachbareffekten und Periodizitäten dienen, die aber mit großer Wahrscheinlichkeit nicht gefaltet sind.

Kapitel 6

Übergreifende zusammenfassende Betrachtung

In der vorgelegten Arbeit sollten mehrere Aspekte der Ähnlichkeit zwischen chemischen Kleinstrukturen und Proteinen untersucht werden. Es zeigte sich, daß der Begriff der Ähnlichkeit je nach untersuchtem Teilgebiet anders aufgefaßt werden muß.

Beim Vergleich von Kleinstrukturen sind strukturelle Merkmale wie das Vorhandensein oder die Abwesenheit von chemischen Gruppen ein häufig angewandtes Kriterium, das durch den Tanimoto-Koeffizienten als Ähnlichkeitsmaß verglichen werden kann. Für die untersuchte Wirksamkeit gegen die Aggregation des Prionproteins ist allerdings ein allgemeines Maß wie der Tanimoto-Koeffizient nicht ausreichend: während zum Beispiel im Falle der Phenothiazine selbst geringe Ähnlichkeiten zu wirksamen Substanzen auf Aktivität schließen läßt, sind selbst Strukturen mit großer Ähnlichkeit zum Steroid Budenosid selbst inaktiv (vgl. Tab. 2.1). Im ersten Fall ist also eine Ähnlichkeit zwischen den untersuchten Substanzen pharmakologisch bedeutsam, während die strukturell eng verwandten Steroide ein weites Spektrum an verschiedenen Wirkungen aufweisen.

Durch eine Kombination von 2D- und 3D-Ähnlichkeitssuchen ist es gelungen, neben Substanzen mit großer chemischer Ähnlichkeit zu den Ausgangssubstanzen auch Wirkstoffe aus neuen chemischen Substanzklassen (*scaffold hopper*) aufzuspüren. Ergebnisse der experimentellen Prüfung dieser Strukturen können als weitere Information in neue Struktursuchen einfließen.

Die auf diese Weise aufgefundenen Wirkstoffe zeigen eine interessante Eigenschaft: sie sind, gemessen am Vorhandensein chemischer Gruppen, unähnlich, ähneln sich aber strukturell. Die Ergebnisse der experimentellen Testung werden Einblick schaffen, welche Art von Ähnlichkeit in diesem Falle pharmakologisch relevant ist.

Ähnlichkeiten zwischen Sequenzmotiven wurden am Beispiel der Wechselwirkung mit dem Importrezeptor peroxisomaler Membranproteine untersucht. Eine positionsspezifische Matrix läßt Aussagen darüber zu, ob eine untersuchte Sequenz an Pex19 bindet oder nicht. Die Matrix kann selbst Peptiden, die sich nicht ähneln, eine Ähnlichkeit bezüglich ihrer Affinität zu Pex19 zuweisen. Dies liegt darin begründet, daß an jeder der neun prediktiven Positionen andere Charakteristika der beteiligten Aminosäuren bedeutungsvoll sind.

Mit Hilfe dieser Matrix konnten Pex19-Bindestellen diverser humaner peroxisomaler Membranproteine vorhergesagt und experimentell bestätigt werden. Dies zeigt, daß die Pex19-Bindung an definierte Targetsignale in peroxisomalen Membranproteinen erfolgt, Pex19 also demnach ein spezifischer Importrezeptor und kein unspezifisch bindendes Chaperon ist. Zum anderen konnte durch diesen Befund die Konservierung der Bindestelle von der Hefe zum Menschen nachgewiesen werden.

Die Bedeutung lokaler Sequenzähnlichkeiten in Proteinen wurde am Beispiel der *cis*-Prolylbindung untersucht. Hier zeigte sich, daß eine Bewertung aufgrund der Charakteristika der Aminosäuren in der Nachbarschaft des Prolylrestes mittels Matrizen nur zu unbefriedigenden Ergebnissen führt. Im Gegensatz zu Ähnlichkeiten zwischen peroxisomalen Targetingsequenzen sind die Gemeinsamkeiten von Proteinen mit *cis*-Prolylbindung offensichtlich schwerer faßbar. Dies ist auch dadurch erklärbar, daß Pex19-Interaktionspartner zur Interaktion mit ihrem Bindungspartner gemeinsame Merkmale aufweisen müssen („Schlüssel-Schloß-Prinzip“ [211]), während dies zur Ausbildung eines gemeinsamen strukturellen Merkmals, hier der *cis*-Prolylbindung, nicht zwingend erforderlich ist.

So ist es durchaus denkbar, daß unterschiedliche Zusammenstellungen von Aminosäuren trotz fehlender Ähnlichkeit gleichermaßen zur Ausbildung von *cis*-Prolylbindungen führen. Ähneln hingegen ein Sequenzmotiv einem der vielen möglichen, zur Ausbildung der *cis*-Peptidylbindung führenden Motive, so ist die Wahrscheinlichkeit gegeben, daß dies ebenfalls eine *cis*-Peptidylbindung ausbildet. Daher war in diesem Falle die Vorhersage mittels Alignments oder Vergleichen zu anderen Sequenzmotiven deutlich erfolgreicher als das Anwenden einer Vorhersagematrix.

Durch die Untersuchung inverser Sequenzähnlichkeiten wurde der Einfluß von scheinbaren Gemeinsamkeiten zwischen Proteinen ohne evolutionären Ursprung evaluiert. Es konnte gezeigt werden, daß eine inverse Sequenzähnlichkeit zwischen Proteinen entgegen früheren Erwartungen keine strukturelle Ähnlichkeit impliziert. Lediglich kleinere Sekundärstrukturelemente, darunter Helices bis zu einer Länge von 49 Aminosäuren, zeigen bei inverser Sequenzähnlichkeit vergleichbare Raumstrukturen.

Als offene Frage verbleibt der Grund für ein gehäuftes Auftreten inverser Sequenzähnlichkeit zwischen Proteinen. Offensichtlich führen Periodizitäten und Häufungen ähnlicher Aminosäuren dazu, daß invertierte Sequenzen „proteinähnlicher“ als randomisierte Sequenzen sind, ohne Rückschlüsse auf Ähnlichkeiten in der Faltung zu erlauben.

Literaturverzeichnis

- [1] VENTER, J. C.; ADAMS, M. D.; MYERS, E. W.; LI, P. W.; MURAL, R. J.; SUTTON, G. G.; SMITH, H. O.; YANDELL, M.; EVANS, C. A.; HOLT, R. A. *et al.*: The sequence of the human genome. *Science* 291 (2001), Nr. 5507, S. 1304–1311
- [2] BLATTNER, F. R.; PLUNKETT, G.; BLOCH, C. A.; PERNA, N. T.; BURLAND, V.; RILEY, M.; COLLADO-VIDES, J.; GLASNER, J. D.; RODE, C. K.; MAYHEW, G. F. *et al.*: The complete genome sequence of *Escherichia coli* K-12. *Science* 277 (1997), Nr. 5331, S. 1453–74
- [3] GOFFEAU, A.; AERT, R.; AGOSTINI-CARBONE, M. L.; AHMED, A.; AIGLE, M.; ALBERGHINA, L.; ALBERMANN, K.; ALBERS, M.; ALDEA, M.; ALEXANDRAKI, D. *et al.*: The Yeast Genome Directory. *Nature* 387 (1997), Nr. suppl, S. 5
- [4] MEWES, H. W.; ALBERMANN, K.; BÄHR, M.; FRISHMAN, D.; GLEISNER, A.; HANI, J.; HEUMANN, K.; KLEINE, K.; MAIERL, A.; OLIVER1, S. G.; PFEIFFER, F.; ZOLLNER, A.: Overview of the yeast genome. *Nature* 387 (1997), Nr. suppl, S. 7–8
- [5] IMANISHI, T.; ITOH, T.; SUZUKI, Y.; O'DONOVAN, C.; FUKUCHI, S.; KOYANAGI, K. O.; BARRERO, R. A.; TAMURA, T.; YAMAGUCHI-KABATA, Y.; TANINO, M. *et al.*: Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biology* 2 (2004), Nr. 6, S. e162
- [6] WESTBROOK, J.; FENG, Z.; CHEN, L.; YANG, H.; BERMAN, H. M.: The Protein Data Bank and structural genomics. *Nucleic Acids Research* 31 (2003), Nr. 1, S. 489–91
- [7] BOECKMANN, B.; BAIROCH, A.; APWEILER, R.; BLATTER, M.-C.; ESTREICHER, A.; GASTEIGER, E.; MARTIN, M.J.; MICHOD, K.;

- O'DONOVAN, C.; PHAN, I.; PILBOUT, S.; SCHNEIDER, M.: The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research* 31 (2003), Nr. 1, S. 365–70
- [8] PRUITT, K. D.; TATUSOVA, T.; MAGLOTT, D. R.: NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* 33 (2005), Nr. 1, S. D501–4
- [9] HEINEMANN, U.; ILLING, G.; OSCHKINAT, H.: High-throughput three-dimensional protein structure determination. *Curr Opin Biotechnol* 12 (2001), Nr. 4, S. 348–54
- [10] ABDEL-RAHMAN, H. M.; AL-KARAMANY, G. S.; EL-KOUSSI, N. A.; YOUSSEF, A. F.; KISO, Y.: HIV protease inhibitors: peptidomimetic drugs and future perspectives. *Curr Med Chem* 9 (2002), Nr. 21, S. 1905–22
- [11] GEORGIADIS, M. M.; JESSEN, S. M.; OGATA, C. M.; TELESNITSKY, A.; GOFF, S. P.; HENDRICKSON, W. A.: Mechanistic implications from the structure of a catalytic fragment of Moloney murine leukemia virus reverse transcriptase. *Structure* 3 (1995), Nr. 9, S. 879–92
- [12] GAO, G.; ORLOVA, M.; GEORGIADIS, M. M.; HENDRICKSON, W. A.; GOFF, S. P.: Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Prot Nat Acad Sci U S A* 94 (1997), Nr. 2, S. 407–11
- [13] BRADLEY, M. P.: An overview of the diversity represented in commercially-available databases. *Mol Divers* 5 (2002), Nr. 4, S. 175–83
- [14] WERMUTH, C. G.: Selective optimization of side activities: another way for drug discovery. *J Med Chem* 47 (2004), Nr. 6, S. 1303–14
- [15] DURANT, J. L.; LELAND, B. A.; HENRY, D. R.; NOURSE, J. G.: Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci* 42 (2002), Nr. 6, S. 1273–80
- [16] DELANEY, J. S.: Assessing the ability of chemical similarity measures to discriminate between active and inactive compounds. *Mol Divers* 1 (1996), Nr. 4, S. 217–22
- [17] MARTIN, Y. C.; KOFRON, J. L.; TRAPHAGEN, L. M.: Do structurally similar molecules have similar biological activity? *J Med Chem* 45 (2002), Nr. 19, S. 4350–8

- [18] WHITTLE, M.; GILLET, V. M.; WILLETT, P.: Enhancing the effectiveness of virtual screening by fusing nearest neighbor lists: A comparison of similarity coefficients. *J Chem Inf Comput Sci* 44 (2004), Nr. 5, S. 1840–48
- [19] KEARSLEY, S. K.: On the orthogonal transformation used for structural comparisons. *Acta Cryst A* 45 (1989), S. 208–10
- [20] THIMM, M.; GOEDE, A.; HOUGARDY, S.; PREISSNER, R.: Comparison of 2D similarity and 3D superposition. Application to searching a conformational drug database. *J Chem Inf Comput Sci* 44 (2004), Nr. 5, S. 1816–22
- [21] COHEN, F. E.; STERNBERG, M. J.: On the prediction of protein structure: The significance of the root-mean-square deviation. *J Mol Biol* 138 (1980), Nr. 2, S. 321–33
- [22] CALZOLAI, L.; LYSEK, D. A.; PEREZ, D. R.; GUNTERT, P.; WUTHRICH, K.: Prion protein NMR structures of chickens, turtles, and frogs. *Prot Nat Acad Sci U S A* 102 (2005), Nr. 3, S. 651–5
- [23] *Kapitel A model of evolutionary change in proteins.* In: DAYHOFF, M. O.; SCHWARTZ, R. M.; ORCUTT, B. C.: *Atlas of protein sequence and structure*. Bd. 5. Washington, D.C. : Natl. Biomed. Res. Found., 1978, S. 345–52
- [24] HENIKOFF, S.; HENIKOFF, J. G.: Amino acid substitution matrices from protein blocks. *Prot Nat Acad Sci U S A* 89 (1992), Nr. 22, S. 10915–9
- [25] NEEDLEMAN, S. B.; WUNSCH, C. D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 48 (1970), Nr. 3, S. 443–53
- [26] SMITH, T.; WATERMAN, M. S.: Identification of common molecular subsequences. *J Mol Biol* 147 (1981), Nr. 1, S. 159–7
- [27] ALTSCHUL, S. F.; GISH, W.; MYERS, W. M.; LIPMAN, D. J.: Basic local alignment search tool. *J Mol Biol* 215 (1990), Nr. 3, S. 403–10
- [28] ALTSCHUL, S. F.; MADDEN, T. L.; SCHAFFER, A. A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. J.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25 (1997), Nr. 17, S. 3389–402

- [29] CHOTHIA, C.; LESK, A. M.: The relation between the divergence of sequence and structure in proteins. *EMBO J* 5 (1986), Nr. 4, S. 823–6
- [30] PREISSNER, R.; GOEDE, A.; MICHALSKI, E.; FRÖMMEL, C.: Inverse sequence similarity in proteins and its relation to the three-dimensional fold. *FEBS Lett* 414 (1997), Nr. 2, S. 425–9
- [31] KENDREW, J. C.; BODO, G.; DINTZIS, H. M.; PARRISH, R. G.; WYCKOFF, H. W.; PHILLIPS, D. C.: A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* 181 (1958), Nr. 4610, S. 662–6
- [32] PERUTZ, M. F.: The first Sir Hans Krebs lecture. X-ray analysis, structure and function of enzymes. *Eur J Biochem* 8 (1969), Nr. 4, S. 445–66
- [33] PAULING, L.; COREY, R. B.; BRANSON, H. R.: The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Prot Nat Acad Sci U S A* 37 (1951), Nr. 4, S. 205–11
- [34] PAULING, L.; COREY, R. B.: The pleated sheet, a new layer configuration of polypeptide chains. *Prot Nat Acad Sci U S A* 37 (1951), Nr. 5, S. 251–6
- [35] GARRET, R. H.; GRISHAM, C. M.: *Biochemistry*. 2nd. Harcourt College Publishing, 1998
- [36] RAMACHANDRAN, G. N.; RAMAKRISHNAN, C.; SASISEKHARAN, V.: Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7 (1963), S. 95–9
- [37] RAMACHANDRAN, G. N.; SASISEKHARAN, V.: Conformation of polypeptides and proteins. *Adv Protein Chem* 23 (1968), S. 283–438
- [38] WEISS, M. S.; JABS, A.; HILGENFELD, R.: Peptide bonds revisited. *Nature Struct Biol* 5 (1998), Nr. 8, S. 676
- [39] FRÖMMEL, C.; PREISSNER, R.: Prediction of prolyl residues in cis-conformation in protein structures on the basis of the amino acid sequence. *FEBS Lett* 277 (1990), Nr. 1–2, S. 159–63
- [40] BALDWIN, R. L.; ROSE, G. D.: Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem Sci* 24 (1999), S. 26–33

- [41] WRIGHT, P. E.; DYSON, J.; LERNER, R. A.: Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding. *Biochemistry* 27 (1988), S. 7167–75
- [42] DYSON, H. J.; MERUTKA, G.; WALTHO, J. P.; LERNER, R. A.; WRIGHT, P. E.: Folding of peptide fragments comprising the complete sequence of protein models for initiation of protein folding I. Myohemerythrin. *J Mol Biol* 226 (1992), S. 795–817
- [43] MINOR, D. L.; KIM, P. S.: Context-dependent secondary structure formation of a designed protein sequence. *Nature* 380 (1996), Nr. 6576, S. 730–4
- [44] DALAL, S.; BALASUBRAMANIAN, S.; REGAN, L.: Protein alchemy: changing beta-sheet into alpha-helix. *Nat Struct Biol* 4 (1997), Nr. 7, S. 548–52
- [45] MEZEI, M.: Chameleon sequences in the PDB. *Protein Eng* 11 (1998), Nr. 6, S. 411–4
- [46] ROST, B.: Twilight zone of protein sequence alignments. *Protein Eng* 12 (1999), Nr. 2, S. 85–94
- [47] GILLE, C.; GOEDE, A.; PREISSNER, R.; ROTHER, K.; FRÖMMEL, C.: Conservation of substructures in proteins: interfaces of secondary structural elements in proteasomal subunits. *J Mol Biol* 299 (2000), Nr. 4, S. 1147–54
- [48] LAURENTS, D. V.; SUBBIAH, S.; LEVITT, M.: Different protein sequences can give rise to highly similar folds through different stabilizing interactions. *Protein Sci* 3 (1994), Nr. 11, S. 1938–44
- [49] HOPPE, A.; FRÖMMEL, C.: NeedleHaystack: a program for the rapid recognition of local structures in large sets of atomic coordinates. *J Appl Cryst* 36 (2003), Nr. 4, S. 1090–7
- [50] SHINDYALOV, I. N.; BOURNE, P. E.: Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* 11 (1998), Nr. 9, S. 739–47
- [51] HOLM, L.; SANDER, C.: Protein structure comparison by alignment of distance matrices. *J Mol Biol* 233 (1993), Nr. 1, S. 123–38
- [52] ANFINSEN, C. B.: Principles that govern the folding of protein chains. *Science* 181 (1973), Nr. 96, S. 223–30

- [53] SELKOE, D. J.: Folding proteins in fatal ways. *Nature* 426 (2003), Nr. 18, S. 900–4
- [54] CHO, H. J.: Is the scrapie agent a virus? *Nature* 262 (1976), Nr. 5567, S. 411–2
- [55] ALPER, T.; CRAMP, W. A.; HAIG, D. A.; CLARKE, M. C.: Does the agent of scrapie replicate without nucleic acid? *Nature* 214 (1967), Nr. 90, S. 764–6
- [56] GRIFFITH, J. S.: Self-replication and scrapie. *Nature* 215 (1967), Nr. 105, S. 1043–4
- [57] PRUSINER, S. B.: Novel proteinaceous infectious particle causes scrapie. *Science* 216 (1982), Nr. 4542, S. 136–44
- [58] COLLINGE, J.: Prion diseases of humans and animals: their causes and molecular bases. *Annu Rev Neurosci* 24 (2001), S. 519–50
- [59] HSIAO, K. K.; SCOTT, M.; FOSTER, D.; GROTH, D. F.; DEARMOND, S. J.; PRUSINER, S. B.: Spontaneous neurodegeneration in transgenic mice with mutant prion protein. *Science* 250 (1990), Nr. 4987, S. 1587–90
- [60] BUELER, H.; AGUZZI, A.; SAILER, A.; GREINER, R. A.; AUTENRIED, P.; AGUET, M.; WEISSMANN, C.: Mice devoid of PrP are resistant to scrapie. *Cell* 73 (1993), Nr. 7, S. 1339–47
- [61] KOCISKO, D. A.; COME, J. H.; PRIOLA, S. A.; CHESEBRO, B.; RAYMOND, G. J.; LANSBURY, P. T.; CAUGHEY, B.: Cell-free formation of protease-resistant prion protein. *Nature* 370 (1994), Nr. 6489, S. 471–4
- [62] SABORIO, G. P.; PERMANNE, B.; SOTO, C.: Sensitive detection of pathological prion protein by cyclic amplification of protein misfolding. *Nature* 411 (2001), Nr. 6839, S. 810–3
- [63] CASTILLA, J.; SAÁ, P.; HETZ, C.; SOTO, C.: In vitro Generation of infectious Scrapie prions. *Cell* 121 (2005), Nr. 2, S. 195–206
- [64] PAN, K. M.; BALDWIN, M.; NGUYEN, J.; GASSET, M.; SERBAN, A.; GROTH, D.; MEHLHORN, I.; HUANG, Z.; FLETTERICK, R. J.; COHEN, F. E.; PRUSINER, S. B.: Conversion of α -helices into β -sheets features in the formation of the scrapie prion proteins. *Proc Natl Acad Sci U S A* 90 (1993), Nr. 23, S. 10962–6

- [65] HUANG, Z.; PRUSINER, S. B.; COHEN, F. E.: Scrapie prions: a three-dimensional model of an infectious fragment. *Fold Des* 1 (1996), Nr. 1, S. 13–19
- [66] CALLAHAN, M. A.; XIONG, L.; CAUGHEY, B.: Reversibility of Scrapie-associated prion protein aggregation. *J Biol Chem* 276 (2001), Nr. 30, S. 28022–8
- [67] CHABRY, J.; CAUGHEY, B.; CHESEBRO, B.: Specific inhibition of *in vitro* formation of protease resistant prion protein by synthetic peptides. *J Biol Chem* 273 (1998), Nr. 21, S. 13203–7
- [68] HORIUCHI, M.; BARON, G. S.; XIONG, L. W.; CAUGHEY, B.: Inhibition of interactions and interconversions of prion protein isoforms by peptide fragments from the C-terminal folded domain. *J Biol Chem* 276 (2001), Nr. 18, S. 15489–97
- [69] HORIUCHI, M.; CAUGHEY, B.: Specific binding of normal prion protein to the scrapie form via a localized domain initiates its conversion to the protease-resistant state. *EMBO J* 18 (1999), Nr. 12, S. 3193–203
- [70] PERETZ, D.; WILLIAMSON, R. A.; KANEKO, K.; VERGARA, J.; LECCLERC, E.; SCHMITT-ULMS, G.; MEHLHORN, I. R.; LEGNAME, G.; WORMALD, M. R.; RUDD, P. M.; DWEK, R. A.; BURTON, D. R.; PRUSINER, S. B.: Antibodies inhibit prion propagation and clear cell cultures of prion infectivity. *Nature* 412 (2001), Nr. 6848, S. 739–43
- [71] BERINGUE, V.; VILETTE, D.; MALLINSON, G.; ARCHER, F.; KAISAR, M.; TAYEBI, M.; JACKSON, G. S.; CLARKE, A. R.; LAUDE, H.; COLLINGE, J.; HAWKE, S.: PrP^{Sc} binding antibodies are potent inhibitors of prion replication in cell lines. *J Biol Chem* 279 (2004), Nr. 38, S. 39671–6
- [72] HEPPNER, F. L.; MUSAHL, C.; ARRIGHI, I.; KLEIN, M. A.; RULICKE, T.; OESCH, B.; ZINKERNAGEL, R. M.; KALINKE, U.; AGUZZI, A.: Prevention of scrapie pathogenesis by transgenic expression of anti-prion protein antibodies. *Science* 294 (2001), Nr. 5540, S. 178–82
- [73] CAUGHEY, W. S.; RAYMOND, L. D.; HORIUCHI, M.; CAUGHEY, B.: Inhibition of protease-resistant prion protein formation by porphyrins and phthalocyanines. *Prot Nat Acad Sci U S A* 95 (1998), Nr. 21, S. 12117–22

- [74] SUPATTAPONE, S.; WILLE, H.; UYECCHI, L.; SAFAR, J.; TREMBLAY, P.; SZOKA, F. C.; COHEN, F. E.; PRUSINER, S. B.; SCOTT, M. R.: Branched polyamines cure prion-infected neuroblastoma cells. *J Virol* 75 (2001), Nr. 7, S. 3453–61
- [75] CAUGHEY, B.; RAYMOND, G. J.: Sulfated polyanion inhibition of scrapie-associated PrP accumulation in cultured cells. *J Virol* 67 (1993), Nr. 2, S. 643–50
- [76] PRIOLA, S. A.; RAINES, A.; CAUGHEY, W. S.: Porphyrin and phthalocyanine antiscrapie compounds. *Science* 287 (2000), Nr. 5457, S. 1503–6
- [77] LADOGANA, A.; CASACCIA, P.; INGROSSO, L.; CIBATI, M.; SALVATORE, M.; XI, Y. G.; MASULLO, C.; POCCHIARI, M.: Sulphate polyanions prolong the incubation period of scrapie-infected hamsters. *J Gen Virol* 73 (1992), Nr. Pt 3, S. 661–5
- [78] DEMAIMAY, R.; HARPER, J.; GORDON, H.; WEAVER, D.; CHESEBRO, B.; CAUGHEY, B.: Structural aspects of Congo red as an inhibitor of protease-resistant prion protein formation. *J Neurochem* 71 (1998), Nr. 6, S. 2534–41
- [79] SELLARAJAH, S.; LEKISHVILI, T.; BOWRING, C.; THOMPSETT, A. R.; RUDYK, H.; BIRKETT, C. R.; BROWN, D. R.; GILBERT, I. H.: Synthesis of analogues of Congo red and evaluation of their anti-prion activity. *J Med Chem* 47 (2004), Nr. 22, S. 5515–34
- [80] CAUGHEY, B.; RAYMOND, L. D.; RAYMOND, G. J.; MAXSON, L.; SILVEIRA, J.; BARON, G. S.: Inhibition of protease-resistant prion protein accumulation in vitro by curcumin. *J Virol* 77 (2003), Nr. 9, S. 5499–502
- [81] KORTH, C.; MAY, B. C.; COHEN, F. E.; PRUSINER, S. B.: Acridine and phenothiazine derivatives as pharmacotherapeutics for prion disease. *Prot Nat Acad Sci U S A* 98 (2001), Nr. 17, S. 9836–41
- [82] BARRET, A.; TAGLIAVINI, F.; FORLONI, G.; BATE, C.; SALMONA, M.; COLOMBO, L.; DE LUIGI, A.; LIMIDO, L.; SUARDI, S.; ROSSI, G.; AUVRE, F.; ADJOU, K. T.; SALES, N.; WILLIAMS, A.; LASMEZAS, C.; DESLYS, J. P.: Evaluation of quinacrine treatment for prion diseases. *J Virol* 77 (2003), Nr. 15, S. 8462–9

- [83] INGROSSO, L.; LADOGANA, A.; POCCHIARI, M.: Congo red prolongs the incubation period in scrapie-infected hamsters. *J Virol* 69 (1995), Nr. 1, S. 506–8
- [84] RUDYK, H.; VASILJEVIC, S.; HENNION, R. M.; BIRKETT, C. R.; HOPE, J.; GILBERT, I. H.: Screening congo red and its analogs for the ability to prevent the formation of PrP^{res} in scrapie-infected cells. *J Gen Virol* 81 (2000), Nr. Pt 4, S. 1155–64
- [85] KLUNK, W. E.; DEBNATH, M. L.; PETTEGREW, J. W.: Development of small molecule probes for the β -amyloid protein of Alzheimer's disease. *Neurobiology of Aging* 15 (1994), Nr. 6, S. 691–8
- [86] COLLINS, S. J.; LEWIS, V.; BRAZIER, M.; HILL, A. F.; FLETCHER, A.; MASTERS, C. L.: Quinacrine does not prolong survival in a murine Creutzfeldt-Jakob disease model. *Ann Neurol* 52 (2002), Nr. 4, S. 503–6
- [87] JOSEFSON, D.: Drugs for malaria and psychosis may offer hope to people with CJD. *British Medical Journal* 323 (2001), Nr. 7310, S. 416
- [88] NAKAJIMA, M.; YAMADA, T.; KUSUHARA, T.; FURUKAWA, H.; TAKAHASHI, M.; YAMAUCHI, A.; KATAOKA, Y.: Results of quinacrine administration to patients with Creutzfeldt-Jakob disease. *Dement Geriatr Cogn Disord* 17 (2004), Nr. 3, S. 158–63
- [89] GILCH, S.; WINKLHOFFER, K. F.; GROSCHUP, M. H.; NUNZIANTE, M.; LUCASSEN, R.; SPIELHAUPT, C.; MURANYI, W.; RIESNER, D.; TATZELT, J.; SCHÄTZL, H. M.: Intracellular re-routing of prion protein prevents propagation of PrP^{Sc} and delays onset of prion disease. *EMBO J* 20 (2001), Nr. 15, S. 3957–66
- [90] KOCISKO, D. A.; BARON, G. S.; RUBENSTEIN, R.; CHEN, J.; KUIZON, S.; CAUGHEY, B.: New inhibitors of scrapie-associated prion protein formation in a library of 2000 drugs and natural products. *J Virol* 77 (2003), Nr. 19, S. 10288–94
- [91] CHANDLER, R. L.: Encephalopathy in mice produced by inoculation with scrapie brain material. *Lancet* 1 (1961), S. 1378–9
- [92] VOIGT, J. H.; BIENFAIT, B.; WANG, S.; NICKLAUS, M. C.: Comparison of the NCI open database with seven large chemical structural databases. *J Chem Inf Comput Sci* 41 (2001), Nr. 3, S. 702–12

- [93] GOEDE, A.; DUNKEL, M.; MESTER, N.; FROMMEL, C.; PREISSNER, R.: SuperDrug: a conformational drug database. *Bioinformatics* in press (2005)
- [94] WHO: The selection and use of essential medicines. Report of the WHO Expert Committee, 2002 (including the 12th Model list of essential medicines). *World Health Organ Tech Rep Ser* 914 (2003), Nr. i-vi, S. 1–126
- [95] MILLER, G. C.; BRITT, H.: A new drug classification for computer systems: the ATC extension code. *Int J Biomed Comput* 40 (1995), Nr. 2, S. 121–4
- [96] DEDUVE, C.; BAUDHUIN, P.: Peroxisomes (microbodies and related particles). *Physiol. Rev.* 46 (1966), Nr. 2, S. 323–57
- [97] KUNAU, W.-H.; BUHNE, S.; GARZA, M. de l.; KIONKA, C.; MATEBLOWSKI, M.; SCHULTZ-BORCHARD, U.; THIERINGER, R.: Comparative enzymology of β -oxidation. *Biochem Soc Trans* 16 (1988), Nr. 3, S. 418–20
- [98] BOSCH, H. van d.; SCHUTGENS, R. B.; WANDERS, R. J.; TAGER, J. M.: Biochemistry of peroxisomes. *Annu Rev Biochem* 61 (1992), S. 157–97
- [99] WANDERS, R. J.; VREKEN, P.; FERDINANDUSSE, S.; JANSEN, G. A.; WATERHAM, H. R.; ROERMUND, C. W.; GRUNSVEN, E. G.: Peroxisomal fatty acid alpha- and beta-oxidation in humans: enzymology, peroxisomal metabolite transporters and peroxisomal diseases. *Biochem Soc Trans* 29 (2001), Nr. Pt 2, S. 250–67
- [100] *Kapitel Peroxisomes in biology and medicine.* In: KUNAU, W.-H.; KIONKA, C.; LEDEBRU, A.; MATEBLOWSKI, M.; GARZA, M. M. l.; SCHULZ-BORCHARD, U.; THIERINGER, R.; VEENHUIS, M.: *β -Oxidation system in eucaryotic microorganisms.* Springer-Verlag, Heidelberg, 1987, S. 128–40
- [101] BIARDI, L.; KRISANS, S. K.: Compartmentalization of cholesterol biosynthesis. Conversion of mevalonate to farnesyl diphosphate occurs in peroxisomes. *J Biol Chem* 271 (1996), Nr. 3, S. 1784–8
- [102] KRISANS, S. K.: The role of peroxisomes in cholesterol metabolism. *Am J Respir Cell Mol Biol* 7 (1992), Nr. 4, S. 358–64

- [103] WANDERS, R. J.; TAGER, J. M.: Lipid metabolism in peroxisomes in relation to human disease. *Mol Aspects Med* 19 (1998), Nr. 2, S. 69–154
- [104] SCHEPERS, L.; CASTEELS, M.; VAMECQ, J.; PARMENTIER, G.; VELDHoven, P van; MANNAERTS, G P.: β -Oxidation of the carboxyl side chain of prostaglandin E2 in rat liver peroxisomes and mitochondria. *J Biol Chem* 263 (1988), Nr. 6, S. 2724–31
- [105] HAYASHI, S.; FUJIWARA, S.; NOGUCHI, T.: Evolution of urate-degrading enzymes in animal peroxisomes. *Cell Biochem Biophys* 32 (2000), Nr. Spring, S. 123–9
- [106] TAKADA, Y.; NOGUCHI, T.: Ureidoglycolate lyase, a new metalloenzyme of peroxisomal urate degradation in marine fish liver. *Biochem J* 235 (1986), Nr. 2, S. 391–7
- [107] MULLER, W. H.; KRIFT, T. P. d.; KROUWER, A. J.; WOSTEN, H. A.; VOORT, L. H. d.; SMAAL, E. B.; VERKLEIJ, A. J.: Localization of the pathway of the penicillin biosynthesis in *Penicillium chrysogenum*. *EMBO J* 10 (1991), Nr. 2, S. 489–95
- [108] FUJIKI, Y.: Peroxisome biogenesis and peroxisome biogenesis disorders. *FEBS Lett* 476 (2005), Nr. 1–2, S. 42–6
- [109] GOULD, S. J.; VALLE, D.: Peroxisome biogenesis disorders. Genetics and Cell biology. *Trends Genet* 16 (2000), Nr. 8, S. 340–5
- [110] WANDERS, R.; WATERHAM, H.: Peroxisomal disorders I: biochemistry and genetics of peroxisome biogenesis disorders. *Clin Genet* 67 (2005), Nr. 2, S. 107–33
- [111] GOLDFISCHER, S.; MOORE, C. L.; JOHNSON, A. B.; SPIRO, A. J.; VALSAMIS, M. P.; WISNIEWSKI, H. K.; RITCH, R. H.; NORTON, W. T.; RAPIN, I.; GARTNER, L. M.: Peroxisomal and mitochondrial defects in the cerebro-hepato-renal syndrome. *Science* 182 (1973), Nr. 107, S. 62–4
- [112] SLAWECKI, M. L.; DODT, G.; STEINBERG, S.; MOSER, A. B.; MOSER, H. W.; GOULD, S. J.: Identification of three distinct peroxisomal protein import defects in patients with peroxisome biogenesis disorders. *J Cell Sci* 108 (1995), Nr. Pt 5, S. 1817–29
- [113] SANTOS, M. J.; IMANAKA, T.; SHIO, H.; SMALL, G. M.; LAZAROW, P. B.: Peroxisomal membrane ghosts in Zellweger syndrome – aberrant organelle assembly. *J Biol Chem* 263 (1988), Nr. 4847, S. 2724–31

- [114] BRAVERMAN, N.; STEEL, G.; OBIE, C.; MOSER, A.; MOSER, H.; GOULD, S. J.; VALLE, D.: Human PEX7 encodes the peroxisomal PTS2 receptor and is responsible for rhizomelic chondrodysplasia punctata. *Nature Genetics* 15 (1997), Nr. 4, S. 369–76
- [115] MOTLEY, A. M.; HETTEMA, E. H.; HOGENHOUT, E. M.; BRITES, P.; ASBROEK, A. L.; WIJBURG, F. A.; BAAS, F.; HEIJMANS, H. S.; TABAK, H. F.; WANDERS, R. J.; DISTEL, B.: Rhizomelic chondrodysplasia punctata is a peroxisomal protein targeting disease caused by a non-functional PTS2 receptor. *Nature Genetics* 15 (1997), Nr. 4, S. 377–80
- [116] PURDUE, P. E.; ZHANG, J. W.; SKONECZNY, M.; LAZAROW, P. B.: Rhizomelic chondrodysplasia punctata is caused by deficiency of human PEX7, a homologue of the yeast PTS2 receptor. *Nature Genetics* 15 (1997), Nr. 4, S. 381–4
- [117] JANSEN, G. A.; OFMAN, R.; FERDINANDUSSE, S.; IJLST, L.; MUIJSERS, A. O.; SKJELDAL, O. H.; STOKKE, O.; JAKOBS, C.; BESLEY, G. T.; WRAITH, J. E.; WANDERS, R. J.: Refsum disease is caused by mutations in the phytanoyl-CoA hydroxylase gene. *Nature Genetics* 17 (1997), Nr. 2, S. 190–3
- [118] LAZAROW, P. B.; FUJIKI, Y.: Biogenesis of peroxisomes. *Annu Rev Cell Biol* 1 (1985), S. 489–530
- [119] GOULD, S. J.; KELLER, G. A.; SCHNEIDER, M.; HOWELL, S. H.; GARRARD, L. J.; GOODMAN, J. M.; DISTEL, B.; TABAK, H.; SUBRAMANI, S.: Peroxisomal protein import is conserved between yeast, plants, insects and mammals. *EMBO J* 9 (1990), Nr. 1, S. 85–90
- [120] HOOP, M. J.; AB, G.: Import of proteins into peroxisomes and other microbodies. *Biochem J* 286 (1992), Nr. Pt 3, S. 657–69
- [121] SUBRAMANI, S.: Protein import into peroxisomes and the biogenesis of the organelle. *Annu Rev Cell Biol* 9 (1993), S. 445–78
- [122] MCCOLLUM, D.; MONOSOV, E.; SUBRAMANI, S.: The pas8 mutant of *Pichia pastoris* exhibits the peroxisomal protein import deficiencies of Zellweger syndrome cells—the PAS8 protein binds to the COOH-terminal tripeptide peroxisomal targeting signal, and is a member of the TPR protein family. *J Cell Biol* 121 (1993), Nr. 4, S. 761–74

- [123] ELGERSMA, Y.; ELGERSMA-HOOISMA, M.; WENZEL, T.; MCCAFFERY, J. M.; FARQUHAR, M. G.; SUBRAMANI, S.: A mobile PTS2 receptor for peroxisomal protein import in *Pichia pastoris*. *J Cell Biol* 140 (1998), Nr. 4, S. 807–20
- [124] REHLING, P.; MARZIOCH, M.; NIESEN, F.; WITTKE, E.; VEENHUIS, M.; KUNAU, W.-H.: The import receptor for the peroxisomal targeting signal 2 (PTS2) in *Saccharomyces cerevisiae* is encoded by the PAS7 gene. *EMBO J* 15 (1996), Nr. 12, S. 2901–13
- [125] NEUBERGER, G.; MAURER-STROH, S.; EISENHABER, B.; HARTIG, A.; EISENHABER, F.: Motif refinement of the peroxisomal targeting signal 1 and evaluation of taxon-specific differences. *J Mol Biol* 328 (2003), Nr. 3, S. 567–79
- [126] REUMANN, S.: Specification of the peroxisome targeting signals type 1 and type 2 of plant peroxisomes by bioinformatics analyses. *Bioinformatics* 135 (2004), Nr. 2, S. 783–800
- [127] NEUBAUER, G.; MAURER-STROH, S.; EISENHABER, B.; HARTIG, A.; EISENHABER, F.: Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. *J Mol Biol* 328 (2003), Nr. 3, S. 581–92
- [128] HETTEMA, E. H.; GIRZALSKY, W.; BERG, M. van d.; ERDMANN, R.; DISTEL, B.: *Saccharomyces cerevisiae* pex3p and pex19p are required for proper localization and stability of peroxisomal membrane proteins. *EMBO J* 19 (2000), Nr. 2, S. 223–33
- [129] HONSHO, M.; TAMURA, S.; SHIMOZAWA, N.; SUZUKI, Y.; KONDO, N.; FUJIKI, Y.: Mutation in PEX16 is causal in the peroxisome-deficient Zellweger syndrome of complementation group D. *Am J Hum Genet* 63 (1998), Nr. 3, S. 1622–30
- [130] GOTTE, K.; GIRZALSKY, W.; LINKERT, M.; BAUMGART, E.; KAMMERER, S.; KUNAU, W.-H.; ERDMANN, R.: Pex19p, a farnesylated protein essential for peroxisome biogenesis. *Mol Cell Biol* 18 (1998), Nr. 1, S. 616–28
- [131] BIERMANN, M.; LAAR, J. von; BROSIUS, U.; GÄRTNER, J.: The peroxisomal membrane targeting elements of human peroxin 2 (PEX2). *Eur J Cell Biol* 82 (2003), Nr. 4, S. 155–62

- [132] SACKSTEDER, K. A.; JONES, J. M.; SOUTH, S. T.; LI, X.; LIU, Y.; GOULD, S. J.: PEX19 binds multiple peroxisomal membrane proteins, is predominantly cytoplasmic, and is required for peroxisome membrane synthesis. *J Cell Biol* 148 (2000), Nr. 5, S. 931–44
- [133] FRANSEN, M.; WYLIN, T.; BREES, C.; MANNAERTS, G. P.; VAN VELDHoven, P. P.: Human pex19p binds peroxisomal integral membrane proteins at regions distinct from their sorting sequences. *Mol Cell Biol* 21 (2001), Nr. 13, S. 4413–24
- [134] JONES, J. M.; MORRELL, J. C.; GOULD, S.: Multiple distinct targeting signals in integral peroxisomal membrane proteins. *J Cell Biol* 153 (2001), Nr. 6, S. 1141–50
- [135] JONES, J. M.; MORRELL, J. C.; GOULD, S. J.: PEX19 is a predominantly cytosolic chaperone and import receptor for class 1 peroxisomal membrane proteins. *J Cell Biol* 164 (2004), Nr. 1, S. 57–67
- [136] HONSHO, M.; HIROSHIGE, T.; FUJIKI, Y.: The membrane biogenesis peroxin Pex16p. Topogenesis and functional roles in peroxisomal membrane assembly. *J Biol Chem* 277 (2002), Nr. 46, S. 44513–24
- [137] BROSIUS, U.; DEHMEL, T.; GÄRTNER, J.: Two different targeting signals direct human peroxisomal membrane protein 22 to peroxisomes. *J Biol Chem* 277 (2002), Nr. 1, S. 774–84
- [138] GLÖCKNER, C. J.; MAYERHOFER, P. U.; LANDGRAF, P.; MUNTAU, A. C.; HOLZINGER, A.; GERBER, J. K.; KAMMERER, S.; ADAMSKI, J.; ROSCHER, A. A.: Human adrenoleukodystrophy protein and related peroxisomal ABC transporters interact with the peroxisomal assembly protein PEX19p. *Biochem Biophys Res Commun* 271 (2000), Nr. 1, S. 144–50
- [139] LANDGRAF, P.; MAYERHOFER, P. U.; POLANETZ, R.; ROSCHER, A. A.; HOLZINGER, A.: Targeting of the human adrenoleukodystrophy protein to the peroxisomal membrane by an internal region containing a highly conserved motif. *Eur J Cell Biol* 82 (2003), Nr. 8, S. 401–10
- [140] BIERMANN, M.; GARTNER, J.: Targeting elements in the amino-terminal part direct the human 70-kDa peroxisomal integral membrane protein (PMP70) to peroxisomes. *Biochem Biophys Res Commun* 285 (2001), Nr. 3, S. 649–55

- [141] FRANSEN, M.; BREES, C.; GHYS, K.; AMERY, L.; MANNAERTS, G. P.; LADANT, D.; VAN VELDHoven, P. P.: Analysis of mammalian peroxin interactions using a non-transcription-based bacterial two-hybrid assay. *Mol Cell Proteomics* 1 (2002), Nr. 3, S. 243–52
- [142] LANDGRAF, C.; PANNI, S.; MONTECCHI-PALAZZI, L.; CASTAGNOLI, L.; SCHNEIDER-MERGENER, J.; VOLKMER-ENGERT, R.; CESARENI, G.: Protein interaction networks by proteome peptide scanning. *PLoS Biology* 2 (2004), Nr. 1, S. E14
- [143] REINEKE, U.; VOLKMER-ENGERT, R.; SCHNEIDER-MERGENER, J.: Applications of peptide arrays prepared by the SPOT-technology. *Curr Opin Biotechnol* 12 (2001), Nr. 1, S. 59–64
- [144] WEISER, A. A.; OR-GUIL, M.; TAPIA, V.; LEICHSENRING, A.; SCHUCHARDT, J.; FRÖMMEL, C.; VOLKMER-ENGERT, R.: SPOT synthesis: Reliability of array-based measurement of peptide binding affinity. *Analytical Biochemistry* in press (2005)
- [145] ROTTENSTEINER, H.; KRAMER, A.; LORENZEN, S.; STEIN, K.; LANDGRAF, C.; VOLKMER-ENGERT, R.; ERDMANN, R.: Peroxisomal Membrane Proteins Contain Common Pex19-binding Sites that are an Integral Part of their Targeting Signals. *Mol Biol Cell* 15 (2004), Nr. 7, S. 3406–17
- [146] PETERS, B.; TONG, W.; SIDNEY, J.; SETTE, A.; WENG, Z.: Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules. *Bioinformatics* 19 (2003), Nr. 14, S. 1765–72
- [147] YU, K. M.; LIU, J.; MOY, R.; LIN, H. C.; NICHOLAS, H. B. J.; ROSENQUIST, G. L.: Prediction of tyrosine sulfation in seven-transmembrane peptide receptors. *Endocrine* 19 (2002), Nr. 3, S. 333–8
- [148] KNELLER, D. G.; COHEN, F. E.; LANGRIDGE, R.: Improvements in protein secondary structure prediction by an enhanced neural network. *J Mol Biol* 214 (1990), Nr. 1, S. 1171–82
- [149] PETERS, B.; SETTE, A.: Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 6 (2005), Nr. 1, S. 132
- [150] HALBACH, A.; LORENZEN, S.; LANDGRAF, C.; VOLKMER-ENGERT, R.; ERDMANN, R.; ROTTENSTEINER, H.: Function of the PEX19-binding site of human ALDP as targeting motif in man and yeast: PMP targeting is evolutionarily conserved. *J Biol Chem* in press (2005)

- [151] MARSHALL, R. D.: The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem Soc Symp* 40 (1974), S. 17–26
- [152] MELLQUIST, J. L.; KASTURI, L.; SPITALNIK, S. L.; SHAKIN-ESHLEMAN, S. H.: The amino acid following an Asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. *Biochemistry* 37 (1998), Nr. 19, S. 6833–7
- [153] HUANG, H. D.; LEE, T. Y.; TZENG, S. W.; WU, L. C.; HORNG, J. T.; TSOU, A. P.; HUANG, K. T.: Incorporating hidden Markov models for identifying protein kinase-specific phosphorylation sites. *J Comput Chem* 26 (2005), Nr. 10, S. 1032–41
- [154] KIM, J. H.; LEE, J.; OH, B.; KIMM, K.; KOH, I.: Prediction of phosphorylation sites using SVMs. *Bioinformatics* 20 (2004), Nr. 17, S. 3179–84
- [155] PODELL, S.; GRIBSKOV, M.: Predicting N-terminal myristoylation sites in plant proteins. *BMC Genomics* 5 (2004), Nr. 1, S. 37
- [156] EISENHABER, B.; EISENHABER, F.; MAURER-STROH, S.; NEUBERGER, G.: Prediction of sequence signals for lipid post-translational modifications: insights from case studies. *Proteomics* 4 (2004), Nr. 6, S. 1614–25
- [157] ROGERS, S.; WELLS, R.; RECHSTEINER, M.: Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* 234 (1986), Nr. 4774, S. 364–8
- [158] ROST, B.: Review: Protein secondary structure prediction continues to rise. *J Struct Biol* 134 (2001), Nr. 2–3, S. 204–18
- [159] BRAUER, A. B. E.; DOMINGO, G. J.; COOKE, R. M.; MATTHEWS, S. J.; LEATHERBARROW, R. J.: A conserved cis peptide bond is necessary for the activity of Bowman-Birk inhibitor protein. *Biochemistry* 41 (2002), Nr. 34, S. 10608–15
- [160] EVANS, T. C.; NELSESTUEN, G. L.: Importance of cis proline 22 in the membrane-binding conformation of bovine prothrombin. *Biochemistry* 35 (1996), Nr. 25, S. 8210–5
- [161] NG, K. K.; WEIS, W. I.: Coupling of prolyl peptide bond isomerization and Ca²⁺ binding in a C-type mannose-binding protein. *Biochemistry* 37 (1998), Nr. 51, S. 17977–89

- [162] LU, K. P.; HANES, S. D.; HUNTER, T.: A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature* 1996 (1996), Nr. 6574, S. 544–7
- [163] SHEN, M.; STUKENBERG, P. T.; KIRSCHNER, M. W.; LU, K. P.: The essential mitotoc peptidyl-prolyl isomerase Pin1 binds and regulates mitosis-specific phosphoproteins. *Genes Dev.* 12 (1998), Nr. 5, S. 706–20
- [164] WU, X.; WILCOX, C. B.; DEVASAHAYAM, G.; HACKETT, R. L.; AREVALO-RODRIGUEZ, M.; CARDENAS, M. E.; HEITMANN, J.; HANES, S. D.: The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. *EMBO J* 19 (2000), Nr. 14, S. 3727–38
- [165] AREVALO-RODRIGUEZ, M.; CARDENAS, M. E.; WU, X.; HANES, S. D.; HEITMAN, J.: Cyclophilin A and Ess1 interact with and regulate silencing by the Sin3-Rpd3 histone deacetylase. *EMBO J* 19 (2000), Nr. 14, S. 3739–49
- [166] LIM, J.; PING LU, K.: Pinning down phosphorylated tau and tauopathies. *Biochim Biophys Acta* 1739 (2005), Nr. 2–3, S. 311–22
- [167] LU, P. J.; WULF, G.; ZHOU, X. Z.; DAVIES, P.; LU, K. P.: The prolyl isomerase Pin1 restores the function of Alzheimer-associated phosphorylated tau protein. *Nature* 399 (1999), Nr. 6738, S. 784–8
- [168] CHARBONNIER, J.-B.; BELIN, P.; MOUTIEZ, M.; STURA, E. A.; QUÉMÉNEUR, E.: On the role of the cis-proline residue in the active site of DsbA. *Protein Science* 8 (1999), Nr. 1, S. 96–105
- [169] MAYR, L. M.; LANDT, O.; HAHN, U.; SCHMID, F. X.: Stability and folding kinetics of ribonuclease T1 are strongly altered by the replacement of cis-proline 39 with alanine. *J Mol Biol* 231 (1993), Nr. 3, S. 897–912
- [170] BIROLO, L.; MALASHKEVICH, V. N.; CAPITANI, G.; DE LUCA, F.; MORETTA, A.; JANSONIUS, J. N.; MARINO, G.: Functional and structural analysis of cis-proline mutants of Escherichia coli aspartate aminotransferase. *Biochemistry* 38 (1999), Nr. 3, S. 905–13
- [171] JIN, L.; STEC, B.; KANTROWITZ, E. R.: A cis-proline to alanine mutant of E. coli aspartate transcarbamoylase: Kinetic studies and three-dimensional crystal structures. *Biochemistry* 39 (2000), Nr. 27, S. 8058–66

- [172] KELLEY, R. F.; RICHARDS, F. M.: Replacement of proline 76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. *Biochemistry* 26 (1987), Nr. 21, S. 6765–74
- [173] KIEFHABER, T.; GRUNERT, H. P.; HAHN, U.; SCHMID, F. X.: Replacement of a cis proline simplifies the mechanism of ribonuclease T1 folding. *Biochemistry* 29 (1990), Nr. 27, S. 6475–80
- [174] KUWAJIMA, K.; OKAYAMA, N.; YAMAMOTO, K.; ISHIHARA, T.; SUGAI, S.: The Pro117 to glycine mutation of staphylococcal nuclease simplifies the unfolding–folding kinetics. *FEBS Lett* 290 (1991), Nr. 1–2, S. 135–8
- [175] STEWART, D. E.; SARKAR, A.; WAMPLER, J. E.: Occurrence and role of cis peptide bonds in protein structures. *J Mol Biol* 214 (1990), Nr. 1, S. 253–60
- [176] PAL, D.; CHAKRABARTI, P.: Cis peptide bonds in proteins: residues involved, their conformations, interactions and locations. *J Mol Biol* 294 (1999), Nr. 1, S. 271–88
- [177] REIMER, U.; SCHERER, G.; DREWELLO, M.; KRUBER, S.; SCHUTKOWSKI, M.; FISCHER, G.: Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *J Mol Biol* 279 (1998), Nr. 2, S. 449–60
- [178] WANG, M.-L.; LI, W.-J.; XU, W.-B.: Support vector machines for prediction of peptidyl prolyl cis/trans isomerization. *J Peptide Res* 63 (2004), Nr. 1, S. 23–28
- [179] WANG, G.; DUNBRACK, R. L.: PISCES: a protein sequence culling server. *Bioinformatics* 19 (2003), Nr. 12, S. 1589–91
- [180] BREADLEY, A. P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30 (1997), S. 1145–59
- [181] DEANE, C. M.; LUMMIS, S. C.: The role and predicted propensity of conserved proline residues in the 5-HT₃ receptor. *J Biol Chem* 276 (2001), Nr. 41, S. 37962–6
- [182] DODGE, C.; SCHNEIDER, R.; SANDER, C.: The HSSP database of protein structure–sequence alignments and family profiles. *Nucleic Acids Research* 26 (1998), Nr. 1, S. 313–5

- [183] LO CONTE, L.; AILEY, B.; HUBBARD, T. J.; BRENNER, S. E.; MURZIN, A. G.; CHOTHIA, C.: SCOP: a structural classification of proteins database. *Nucleic Acids Research* 28 (2000), Nr. 1, S. 257–9
- [184] MACARTHUR, M. W.; THORNTON, J. M.: Influence of proline residues on protein conformation. *J Mol Biol* 218 (1991), Nr. 2, S. 397–412
- [185] WEISSHOFF, H.; FROST, K.; BRANDT, W.; HENKLEIN, P.; MÜGGE, C.; FRÖMMEL, C.: Novel disulfide-constrained pentapeptides as models for β VIa turns in proteins. *FEBS Lett* 372 (1995), Nr. 2–3, S. 203–9
- [186] UCHIDA, T.; FUJIMORI, F.; TRADLER, T.; FISCHER, G.; RAHFELD, J. U.: Identification and characterization of a 14 kDa human protein as a novel parvulin-like peptidyl prolyl cis/trans isomerase. *FEBS Lett* 446 (1999), Nr. 2–3, S. 278–82
- [187] RAHFELD, J. U.; SCHIERHORN, A.; MANN, K.; FISCHER, G.: A novel peptidyl-prolyl cis/trans isomerase from *Escherichia coli*. *FEBS Lett* 343 (1994), Nr. 1, S. 65–9
- [188] PATZELT, H.; RUDIGER, S.; BREHMER, D.; KRAMER, G.; VORDERWULBECKE, S.; SCHAFFITZEL, E.; WAITZ, A.; HESTERKAMP, T.; DONG, L.; SCHNEIDER-MERGENER, J.; BUKAU, B.; DEUERLING, E.: Binding specificity of *Escherichia coli* trigger factor. *Prot Nat Acad Sci U S A* 98 (2001), Nr. 25, S. 14244–9
- [189] HARRISON, R. K.; STEIN, R. L.: Mechanistic studies of enzymatic and nonenzymatic prolyl cis-trans isomerization. *J Am Chem Soc* 114 (1992), S. 3464–71
- [190] SALI, A.; OVERINGTON, J. P.: Derivation of rules for comparative protein modelling from a database of protein structure alignments. *Protein Science* 234 (1994), Nr. 3, S. 1582–96
- [191] FORSTER, M. J.: Molecular modelling in structural biology. *Micron* 33 (2002), Nr. 4, S. 365–84
- [192] GUPTASARMA, P.: Reversal of peptide backbone direction may result in the mirroring of protein structure. *FEBS Lett* 310 (1992), Nr. 3, S. 205–10
- [193] OLSZEWSKI, K. A.; KOLINSKI, A.; SKOLNICK, J.: Does a backwardly read protein sequence have a unique native state? *Protein Eng* 9 (1996), Nr. 1, S. 5–14

- [194] LACROIX, E.; VIGUERA, A. R.; SERRANO, L.: Reading protein sequences backwards. *Fold Des* 3 (1998), Nr. 2, S. 79–85
- [195] WITTE, K.; SKOLNICK, J.; WONG, C.-H.: A Synthetic Retrotransition (Backward Reading) Sequence of the Right-Handed Three-Helix Bundle Domain (10-53) of Protein A Shows Similarity in Conformation as Predicted by Computation. *J Am Chem Soc* 120 (1998), Nr. 50, S. 13042–5
- [196] LIU, N.; DEILLON, C.; KLAUSER, S.; GUTTE, B.; THOMAS, R. M.: Synthesis, physicochemical characterization, and crystallization of a putative retro-coiled coil. *Protein Science* 7 (1998), Nr. 5, S. 1214–20
- [197] MITTL, P. R.; DEILLON, C.; SARGENT, D.; LIU, N.; KLAUSER, S.; THOMAS, R. M.; GUTTE, B.; GRUTTER, M. G.: The retro-GCN4 leucine zipper sequence forms a stable three-dimensional structure. *Prot Nat Acad Sci U S A* 97 (2000), Nr. 6, S. 2562–6
- [198] CHELEY, S.; BRAHA, O.; LU, X.; CONLAN, S.; BAYLEY, H.: A functional protein pore with a ‘retro’ transmembrane domain. *Protein Science* 8 (1999), Nr. 6, S. 1257–67
- [199] PARK, J.; DIETMANN, S.; HEGER, A.; HOLM, L.: Estimating the significance of sequence order in protein secondary structure and prediction. *Bioinformatics* 16 (2000), Nr. 11, S. 978–87
- [200] CHOREV, M.; GOODMAN, M.: A dozen years of retro-inverso peptidomimetics. *Acc Chem Res* 26 (1993), Nr. 5, S. 266–73
- [201] FLETCHER, M. D.; CAMPBELL, M. M.: Partially Modified Retro-Inverso Peptides: Development, Synthesis, and Conformational Behavior. *Chem Rev* 98 (1998), Nr. 2, S. 763–96
- [202] JUVVADI, P.; VUNNAM, S.; MERRIFIELD, R. B.: Synthetic Melittin, Its Enantio, Retro, and Retroenantio Isomers, and Selected Chimeric Analogs: Their Antibacterial, Hemolytic, and Lipid Bilayer Action. *J Am Chem Soc* 118 (1996), Nr. 38, S. 8989–97
- [203] PHAN-CHAN-DU, A.; PETIT, M. C.; GUICHARD, G.; BRIAND, J. P.; MULLER, S.; CUNG, M. T.: Structure of antibody-bound peptides and retro-inverso analogues. A transferred nuclear Overhauser effect spectroscopy and molecular dynamics approach. *Biochemistry* 40 (2001), Nr. 19, S. 5720–7

- [204] CHOREV, M.; GOODMAN, M.: Recent developments in retro peptides and proteins—an ongoing topochemical exploration. *Trends Biotechnol* 13 (1995), Nr. 10, S. 438–45
- [205] C., Milton R.; C., Milton S.; B., Kent S.: Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity. *Science* 256 (1992), Nr. 5062, S. 1445–8
- [206] HENIKOFF, S.; HENIKOFF, J. G.: Amino acid substitution matrices. *Adv Protein Chem* 54 (2000), S. 73–97
- [207] ANDREEVA, A.; HOWORTH, D.; BRENNER, S. E.; HUBBARD, T. J.; CHOTHIA, C.; MURZIN, A. G.: SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Research* 32 (2004), Nr. Database issue, S. D226–9
- [208] BRENNER, S. E.; KOEHL, P.; LEVITT, M.: The ASTRAL compendium for protein structure and sequence analysis. *Nucleic Acids Research* 28 (2000), Nr. 1, S. 254–6
- [209] PREISSNER, R.; GOEDE, A.; FROMMEL, C.: Spare parts for helix-helix interaction. *Protein Eng* 12 (1999), Nr. 10, S. 825–32
- [210] CARTER, P.; ANDERSEN, C. A.; ROST, B.: DSSPcont: Continuous secondary structure assignments for proteins. *Nucleic Acids Research* 31 (2003), Nr. 13, S. 3293–5
- [211] FISCHER, E.: Einfluss der Configuration auf die Wirkung der Enzyme. *Chem. Ber.* 27 (1894), S. 2985–2993

Eigene Publikationen

S. LORENZEN, C. GILLE, R. PREISSNER, AND C. FRÖMMEL, *Inverse sequence similarity of proteins does not imply structural similarity*, FEBS Lett. 19, pp. 105–9 (2003).

C. GILLE, S. LORENZEN, E. MICHALSKY, AND C. FRÖMMEL, *KISS for STRAP: user extensions for a protein alignment editor*, Bioinformatics 19, pp. 2489–91 (2003).

H. ROTTENSTEINER, A. KRAMER, S. LORENZEN, K. STEIN, C. LANDGRAF, R. VOLKMER-ENGERT, AND R. ERDMANN, *Peroxisomal membrane proteins contain common Pex19p-binding sites that are an integral part of their targeting signals*, Mol Biol Cell. 15, pp. 3406–17 (2004).

S. LORENZEN, B. PETERS, A. GOEDE, R. PREISSNER, AND C. FRÖMMEL, *Conservation of cis prolyl bonds in proteins during evolution*, Proteins 58, pp. 589–95 (2004).

S. LORENZEN, M. DUNKEL, AND R. PREISSNER, *In silico screening of drug databases for TSE inhibitors*, BioSystems 80, pp. 117–22 (2005).

A. HALBACH, S. LORENZEN¹, C. LANDGRAF, R. VOLKMER-ENGERT, R. ERDMANN, AND H. ROTTENSTEINER *PEX19-binding sites are targeting motifs for peroxisomal membrane proteins that are conserved from yeast to man*, J Biol Chem 280, pp. 21176–82 (2005).

Submitted: C. YERNAUX, M. FRANSEN, C. BREES, S. LORENZEN, AND P. A. M. MICHELS *Characterization of ABC transporters in the membrane glycosomes of Trypanosoma brucei*

In Vorbereitung: P. HILDEBRAND, S. LORENZEN, R. PREISSNER, AND C. FRÖMMEL *Prediction of helix – helix interactions in transmembrane channels and transporters*

¹gleichberechtigter Erstautor

Danksagung

Die Erstellung dieser Arbeit wäre ohne die Mithilfe zahlreicher Menschen nicht möglich gewesen. Ihnen möchte ich an dieser Stelle meinen Dank aussprechen.

Zunächst danke ich Prof. Cornelius Frömmel, der mir die Gelegenheit gab, in seiner Arbeitsgruppe an interessanten Themen zu arbeiten.

Robert Preißner lieferte besonders in der Anfangszeit des Projektes viele Ideen und stand mir stets mit Rat und Tat zur Seite.

Experimentelle Daten über Pex19-Bindung wurden im Labor von Prof. Ralf Erdmann von Hanspeter Rottensteiner und André Halbach generiert, die Peptid-Blots stammen von Rudolf Volkmer-Engert und Christiane Landgraf. Die Zusammenarbeit mit diesen beiden Gruppen ließ keine Wünsche offen.

Kristian Rother gilt mein Dank für kritische Anmerkungen zum Manuskript, und Mathias Dunkel und Stefan Günther verdanke ich so manchen lustigen Abend. Melanie Füllbeck, Ines Jäger, Elke Michalsky und in gewissem Maße auch Olga Viebig sorgten stets für eine fesselnde Arbeitsatmosphäre im „Praktikantenraum“.

Christoph Gille stand mir häufig in Sachen Java zur Seite, und ihm, Andreas Hoppe und Thorsten Pöschel ist zu verdanken, daß meine Begeisterung für \LaTeX geweckt wurde. Björn Peters unterwies mich in der hohen Kunst der Vorhersage durch stabilisierte Matrizen, und heiße wissenschaftliche Streitgespräche bei „Können wir Ihnen helfen“ werde ich bestimmt noch lange im Gedächtnis behalten.

Besonders möchte ich an dieser Stelle meiner wunderbaren Freundin Huyên danken, die stets für mich da war und auch in schwierigen Zeiten an meiner Seite stand.

Abschließend danke ich meinen Eltern, ohne deren stete Unterstützung und Hilfe ich es nie so weit geschafft hätte.

Lebenslauf

Name:	Stephan Lorenzen
Geburtsdatum:	23.März 1976
Geburtsort:	Flensburg
Nationalität:	deutsch
Familienstand:	ledig
08/1982 - 07/1986	Grundschule, Grund- und Hauptschulzentrum Schafflund
08/1986 - 07/1995	Gymnasium, Auguste-Viktoria-Schule Flensburg
08/1995 - 07/1996	Wehrdienst
10/1996 - 09/2001	Studium an der Freien Universität zu Berlin in der Fachrichtung Biochemie
02/1999 - 03/1999	Wissenschaftliche Mitarbeit am ZMBH Heidelberg, AG Prof. Dobberstein
05/1999 - 08/1999	Wissenschaftliche Mitarbeit an der Harvard Medical School, AG Prof. Rapoport
02/2001 - 08/2001	Diplomarbeit an der FU Berlin, AG Prof. Ralf Erdmann „Tag-Markierung zur Aufreinigung von Proteinen des peroxisomalen Dockingkomplexes aus <i>S. cerevisiae</i> “
10/2001 - heute	Wissenschaftlicher Mitarbeiter an der Charité, Institut für Biochemie, AG Prof. Frömmel
10/2004	Teilnahme am 4. Kongress der Nationalen TSE-Forschungsplattform, Poster „In silico screening of databases for TSE inhibitors“

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig ohne fremde Hilfe verfaßt und nur die angegebene Literatur und Hilfsmittel verwendet zu haben.

Stephan Lorenzen
30. Juni 2005